1-1-2005

Lessons in Machine Ethics from the Perspective of Two Computational Models of Ethical Reasoning

Bruce M. McLaren
Carnegie Mellon University

Presented in November, 2005 at the AAAI Fall Symposium, Washington, D.C.
In "Papers from the AAAI Fall Symposium", Technical Report FS-05-06, pp. 70-77

Lessons in Machine Ethics from the Perspective of Two Computational Models of Ethical Reasoning

Bruce M. McLaren

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3891
bmclaren@cs.cmu.edu

Abstract

In this paper, two computational models of ethical reasoning, one that compares pairs of truth-telling cases and one that retrieves relevant past cases and principles when presented with an ethical dilemma, are described and discussed. Lessons learned from developing and experimenting with the two systems, as well as challenges of building programs that reason about ethics, are discussed. Finally, plans for developing an intelligent tutor for ethics using one of the computational models as a basis is presented.

Introduction

How can machines support humans in ethical reasoning? This is a question of great interest to those engaged in Machine Ethics research. During the past 15 years, several Artificial Intelligence (AI) programs have been developed to address, or at least begin to address, this question. This paper discusses two of those programs, both developed by the author. One of the programs, Truth-Teller, is designed to accept a pair of ethical dilemmas and describe the salient similarities and differences between the cases, from both an ethical and pragmatic perspective. The other program, SIROCCO, is constructed to accept a single ethical dilemma and retrieve other cases and ethical principles that may be relevant to the new case.

Neither program was designed to reach an ethical decision. The view that runs throughout the author's work is that reaching an ethical conclusion is, in the end, the obligation of a *human* decision maker. Even if the author believed the computational models presented in this paper were up to the task of autonomously reaching correct conclusions to ethical dilemmas, having a computer program propose decisions oversimplifies the obligations of human beings and makes assumptions about the "best" form of ethical reasoning. Rather, the aim in this work has been to develop programs that produce relevant information that can help humans as they struggle with difficult ethical decisions, as opposed to providing fully supported ethical arguments and conclusions. In other words, if the programs can stimulate the "moral imagination" (Harris, Pritchard, and Rabins, p. 19, 2004) and help humans reach decisions, they will have succeeded.

The paper is organized as follows. First, Truth-Teller and SIROCCO are briefly described and compared. Second, lessons learned from the two projects are presented. Finally, the author's current work in the area of Machine Ethics is briefly described: using case comparisons as the basis of an intelligent tutor for ethics.

Truth-Teller

Truth-Teller, the first program implemented by the author to perform ethical reasoning, compares pairs of cases presenting ethical dilemmas about whether or not to tell the truth (Ashley and McLaren, 1994; 1995; McLaren and Ashley, 1995a; 1995b). The program was intended to be a first step in implementing a computational model of casuistic reasoning, a form of ethical reasoning from antiquity in which decisions are made by comparing a problem to paradigmatic, real, or hypothetical cases (Jonsen and Toulmin, 1988). Casuistry long ago fell out of favor with philosophers and ethicists but has recently been employed, in practical dilemmas, by medical ethicists (Strong, 1988; Arras, 1991; Brody, 2003).

The program marshals ethically relevant similarities and differences between two given cases from the perspective of the "truth teller" (i.e., the person faced with the dilemma) and reports them to the user. In particular, it points out reasons for telling the truth (or not) that (1)

apply to both cases, (2) apply more strongly in one case than another or (3) apply to only one case. An example of a comparison made by Truth-Teller is shown below (Figure 1).

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

Truth-Teller has a set of methods for reasoning that enables it to integrate reasons, principles, and cases intelligently in its case comparisons. Broadly characterized, Truth-Teller's methods comprise three phases of analysis for (1) aligning, (2) qualifying, and (3) marshaling reasons, followed by (4) an interpretation phase. Each of the phases is described in more detail below:

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 Qualification Phase. Qualifying a reason means identifying special relationships among actors, actions, and reasons that augment or diminish the importance of the reasons. The qualification phase adjusts the relative importance of competing reasons or principles in the problem. During the qualification phase, heuristic production rules qualify or "tag" objects and the alignments between objects in a variety of ways based on considerations like criticalness, altruism, participants' roles and alternative actions.

The Marshaling Phase. Marshaling reasons means selecting particular similar or differentiating reasons to emphasize in presenting an argument that (1) one case is as strong as or 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. The marshaling phase analyzes the aligned and qualified comparison data, determines how the cases should be compared to one another based on five pre-defined comparison contexts reflecting a qualitative assessment of the overall similarity between the two cases, and then organizes information appropriate to that type of comparison.

The Interpretation Phase. A fourth phase of the program generates the comparison text by interpreting the activities of the first three phases.

Truth-Teller employs two abstraction hierarchies to help it produce its output: a Reasons Hierarchy, which organizes reasons or rationales according to facets that are important in telling the truth (Bok, 1989), and a Relations Hierarchy, which represents human relations (e.g., spouse, friend, business associate) and the incumbent level of duty and trust expected in such relations (Aristotle, edited and published in 1924, Books VIII and IX; Jonsen and Toulmin, 1988, p. 290-293). While specific reasons are used to represent individual cases (e.g., produce benefit for professional status), these reasons are sub-types of more general reasons, such as beneficence, non-maleficence, or justice (Ross, 1930). The two hierarchies are used to classify and moderate the support for and against telling the truth in each scenario. The hierarchies also help Truth-

Teller compare the cases by matching reasons across the cases.

In Truth-Teller, each case is focused on the main protagonist's reasons for and against telling the truth. For instance, Figure 2 depicts Truth-Teller's representation of the Stephanie case of Figure 1. In this case, Stephanie is the "truth teller" and the actor(s) who may receive the truth, i.e., the "truth receivers," are the experiment subjects. Stephanie can take several possible actions: tell the experiment subjects the truth, tell them a lie, or perhaps think of a compromise solution (e.g., partially reveal the truth). Each of these possible actions has reasons that support it. For instance, two reasons for Stephanie to tell the truth are (1) the subjects have the right not to be deceived and (2) Stephanie may be professionally harmed if she is caught lying. Truth-Teller's task is to compare pairs of cases by aligning and comparing the reasons represented in each case.

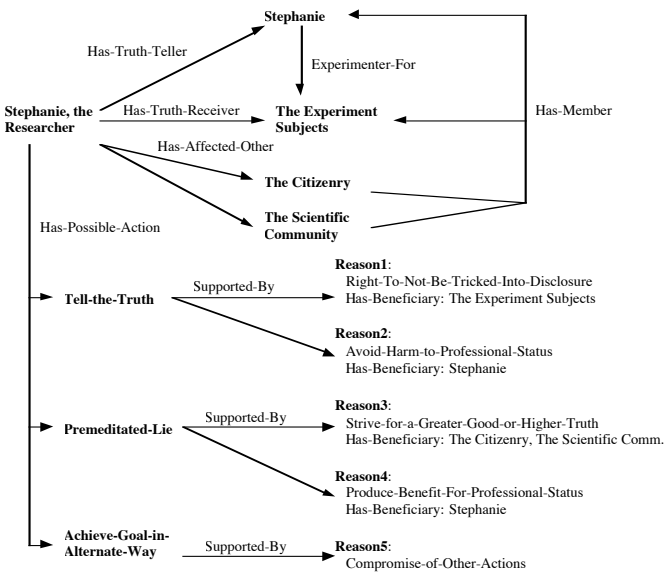

Figure 2: An Example of Truth-Teller's Case Representation

The smallest representational element in Truth-Teller is a *reason*. To represent a case, case enterers must elaborate the ethical and pragmatic reasons for (and against) telling the truth by interpreting the text of the case. Such a representation, while quite useful for generating the detailed and issue-focused comparison texts produced by the program, is constrained in its general applicability. Essentially, Truth-Teller is very good at comparing truth-telling dilemmas in a sophisticated and meaningful way, but it cannot tackle other types of ethical problems without augmentation of its case representation and Reasons Hierarchy.

To test Truth-Teller's ability to compare cases, an evaluation was performed in which professional ethicists were asked to grade the program's output. The goal was to test whether Truth-Teller's case comparisons would be regarded by expert ethicists as high quality, and this was achieved by polling the opinions of five professional ethicists as to the reasonableness (R), completeness (C), and context sensitivity (CS) on a scale of 1 (low) to 10 (high) of twenty of Truth-Teller's case comparisons, similar to and including the comparison in Figure 1. The mean scores assigned by the five experts across the twenty comparisons were R=6.3, C=6.2, and CS=6.1. Two human comparisons, written by post-graduate humans, were also included in the evaluation and, not surprisingly, these comparisons were graded somewhat higher by the ethicists, at mean scores of R=8.2, C=7.7, and CS=7.8. On the other hand, two of Truth-Teller's comparisons graded higher than one of the human evaluations.

These results indicate that Truth-Teller is at least moderately successful at comparing truth-telling dilemmas. Since the expert ethicists were given the instruction to "evaluate comparisons as you would evaluate short answers written by college undergraduates," it is quite encouraging that Truth-Teller performed as well as it did.

SIROCCO

SIROCCO, the second ethical reasoning program created by the author, was developed to explore and analyze the relationship between general principles and concrete facts of cases. In particular, the program was designed to emulate the way in which an ethical review board within a professional engineering organization (the National Society of Professional Engineers – NSPE) decides cases by referring to, and balancing between, ethical codes and past cases. The principles in engineering ethics, while more specific than general duties such as justice and beneficence, still tend to be too general to decide cases, so the NSPE board often uses past cases as precedent in deciding new cases.

SIROCCO's goal, given a new case to analyze, is to provide the basic information with which a human reasoner, for instance a member of the NSPE review board, could answer an ethical question and then build an argument or rationale for that conclusion (McLaren and Ashley, 2000; McLaren, 2003). The program utilizes knowledge of past case analyses, including past retrieval of principles and cases, and the way these knowledge elements were utilized in the past analyses, to support its retrieval and analysis in the new case. The techniques applied by SIROCCO are known as *operationalization techniques*.

An example of SIROCCO's output is shown in Figure 3. The facts of the input case and the question raised by the case are first displayed. This particular case involves an engineer who discovers serious safety problems in a building but does not report the safety problems to anyone except the client, because his client, the building owner, requests confidentiality. The question raised is whether it was ethical for the engineer to give preference to the client's confidentiality over the public's safety.

SIROCCO's output consists of various information, derived by the application of the operationalization techniques, that could support a human in reasoning about and arguing this case: (1) a list of possibly relevant codes, (2) a list of possibly relevant cases, and (3) a list of additional suggestions. Although not illustrated in Figure 3, SIROCCO is also capable of explaining its output: it can display its reasons for selecting the suggested codes and cases. The interested reader can run the SIROCCO program on over 200 ethical dilemmas and view information such as that shown in Figure 3 by going to the following web page:
http://sirocco.lrdc.pitt.edu/sirocco/index.html

****
*** **SIROCCO is analyzing Case 90-5-1**
****

Facts:
Tenants of an apartment building sue the owner to force him to repair many defects in the building that affect the quality of use. The owner's attorney hires Engineer A to inspect the building and give expert testimony in support of the owner. Engineer A discovers serious structural defects in the building, which he believes constitute an immediate threat to the safety of the tenants. The tenants' suit has not mentioned these safety-related defects. Upon reporting the findings to the attorney, Engineer A is told he must maintain this information as confidential as it is part of a lawsuit. Engineer A complies with the request of the attorney.

Question:
Was it ethical for Engineer A to conceal his knowledge of the safety-related defects in view of the fact that it was an attorney who told him he was legally bound to maintain confidentiality?

****
*** SIROCCO has the following suggestions
*** for evaluating '90-5-1: Failure to Report
*** Information Affecting Public Safety"
****

*** ***Possibly Relevant Codes:***

I-4: Act as a Faithful Agent or Trustee
III-4: Do not Disclose Confidential Information Without Consent
I-1: Safety, Health, and Welfare of Public is Paramount
II-1-A: Primary Obligation is to Protect Public (Notify Authority if Judgment is Overruled).
III-1-B: Advise Client or Employer When a Project Will Not Be Successful
III-1: Be Guided by Highest Standards of Integrity
II-1-C: Do not Reveal Confidential Information Without Consent
III-2-B: Do not Complete or Sign Documents that are not Safe for Public
II-1-E: Report Alleged Code Violations
II-5-A: Do not Falsify or Misrepresent Qualifications

*** ***Possibly Relevant Cases:***

76-4-1: Public Welfare - Knowledge of Information Damaging to Client's Interest
89-7-1: Duty To Report Safety Violations
84-5-1: Engineer's Recommendation For Full-Time, On-Site Project Representative

*** ***Additional Suggestions:***

- The codes II-1-A ('Primary Obligation is to Protect Public (Notify Authority if Judgment is Overruled).') and I-1 ('Safety, Health, and Welfare of Public is Paramount') may override codes III-4 ('Do not Disclose Confidential Information Without Consent'), I-4 ('Act as a Faithful Agent or Trustee'), and III-1 ('Be Guided by Highest Standards of Integrity') in this case. See case 76-4-1 for an example of this type of code conflict and resolution.

...

- The case 67-10-1 was cited by 76-4-1 to highlight or elaborate a general principle or common scenario. Since 76-4-1 has been suggested as possibly relevant to the present case, its cited case may also be relevant. Check whether the general scenario of the cited case is relevant to the present case: 'Engineer is involved in a professional situation in which the public welfare is at stake'

Figure 3: An Excerpt of SIROCCO's Output for Case 90-5-1

SIROCCO accepts input, or *target*, cases in a detailed case-representation language called the *Engineering Transcription Language* (ETL). SIROCCO's language represents the actions and events of a scenario as a *Fact Chronology* of individual sentences (i.e., *Facts*), each consisting of (1) *Actors and objects*, instances of general actors and objects which appear in the scenario, (2) a *Fact Primitive,* the action or event in which the actor and/or object instances participated, and (3) a *Time Qualifier,* a temporal relation that specifies how a Fact relates to other Facts in time. A predefined ontology of Actor, Object, Fact Primitive, and Time Qualifier types are used in the representation. At least one Fact in the Fact Chronology is designated as the *Questioned Fact*; this is the action or event corresponding to the ethical question raised in the scenario. The entire ontology, a detailed description of how cases are represented, and over 50 example *Fact Chronologies* can be found at:
http://www.pitt.edu/~bmclaren/ethics/index.html.

SIROCCO employs a two-stage graph-mapping algorithm to retrieve cases and codes, as depicted in Figure 4. Stage 1 performs a "surface match" by retrieving all *source* cases – the cases in the program's database, represented in an extended version of ETL (EETL),

totaling over 400 – that share any fact with the target case. It computes a weighted dot product with all retrieved cases, based on fact matching between the target case and each source case, and outputs a list of candidate source cases ranked by dot product scores. Different weights are assigned to matches at four abstraction levels (i.e., the lowest level matches are weighted higher than more abstract matches). Higher weights are also assigned to matches to critical facts of the source cases.

Using a heuristic A* search, Stage 2 attempts a structural mapping between the target case and each of the N top-ranking candidate source cases from Stage 1. SIROCCO takes temporal relations and abstract matches into account in this search. The search focuses on matching facts from source cases that were relevant to past application of principles and cases (which are represented in the source cases). This focus allows SIROCCO's A* search to be both tractable and more likely to identify salient similarities in past cases. The top-rated structural mappings uncovered by the A* search are organized and displayed by a module called the Analyzer.

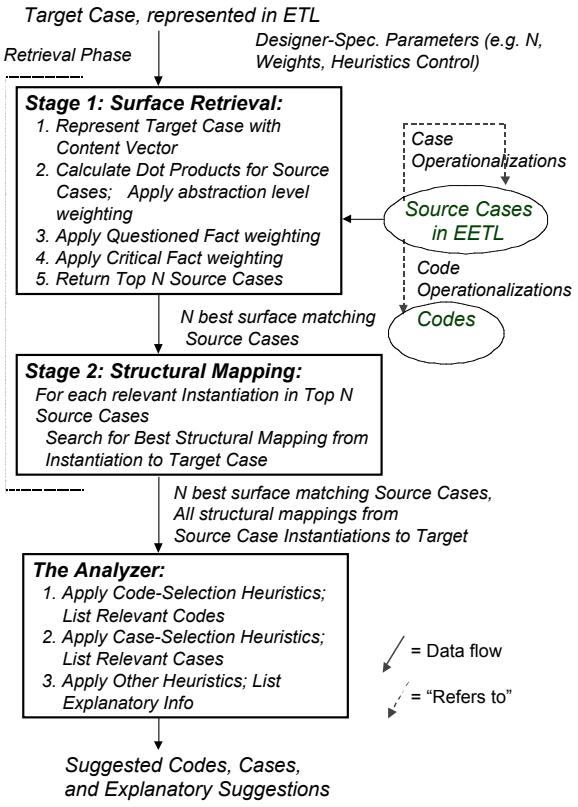

Figure 4: SIROCCO's Algorithm

A large-scale, formal experiment was performed with SIROCCO to test how well it retrieved principles and cases in comparison to several other retrieval techniques, including two full-text retrieval systems (MG and Extended-MG) and a version of SIROCCO that does not use the operationalization techniques. Each method was scored based on how well its retrieved cases and codes overlapped with that of the humans' (i.e., the NSPE review board) retrieved cases and codes in evaluating the same cases, using a metric called the *F-Measure* (Lewis *et al*, 1996). The methods were compared on two dimensions: exact matching (defined as the method and humans retrieving precisely the same codes and cases) and inexact matching (defined as the method and humans retrieving closely related codes and cases). In these experiments, the probability that SIROCCO was more accurate than the other five methods was greater than 95% in every instance except with respect to EXTENDED-MG on the inexact matching. There the probability was 94.3.

Comparison of Truth-Teller and SIROCCO

Fundamentally, Truth-Teller and SIROCCO have different purposes. Truth-Teller is more useful in helping users recognize important similarities and differences between cases. Its output is focused on *explaining* the salient similarities and differences. In fact, as described later in this paper, current work by the author involves tutoring students on how to compare cases using the Truth-Teller approach. While SIROCCO also compares cases, its results are not focused on case comparisons. Rather, SIROCCO is more useful for collecting a variety of information, principles, cases, and additional information that a user should consider in evaluating a new ethical dilemma.

While Truth-Teller has a clear advantage in comparing cases and explaining those comparisons, it ignores the problem of how potentially "comparable" cases are identified in the first place. The program compares any pair of cases it is provided, no matter how different they may be. SIROCCO, on the other hand, uses retrieval to determine which cases are most likely to be relevant to a given target case and thus worth comparing. An interesting synthesis of the two programs would be to have SIROCCO do the work of retrieving comparable cases and Truth-Teller do the work of comparing cases.

To achieve such an integration, however, the two programs would need more common representational elements. In SIROCCO, primitives that closely model some of the actions and events of a fact situation are used to represent cases as complex narratives. In this sense, SIROCCO's representational approach is more sophisticated and general than Truth-Teller's. This is key to SIROCCO's ability to address a wider range of cases than Truth-Teller addresses; not only does SIROCCO handle ethical issues regarding honesty, it can also handle scenarios regarding public safety, confidentiality, conflict of interest and many more. In addition, SIROCCO's representation is more appropriate for untrained case enterers to transcribe cases – it requires far less abstraction from the actual facts of the case and thus enables the collection of a greater number and range of cases. On the other hand, SIROCCO's case comparisons are not nearly

as precise and issue-oriented as Truth-Teller's. This is the trade-off for addressing a wider variety of cases.

Lessons Learned

The first and foremost lesson learned from the Truth-Teller and SIROCCO projects is that ethical reasoning has a fundamentally different character than reasoning in more structured and formalized domains. In ethical reasoning, "inference rules" are available almost exclusively at an abstract level, in the form of principles. The difficulty in addressing and forming arguments in such domains using formal logic has long been recognized (Toulmin, 1958), and some practitioners in Artificial Intelligence, particularly those interested in legal reasoning, have also grappled with this issue. As pointed out by Ashley, "The legal domain is harder to model than mathematical or scientific domains because deductive logic, one of the computer scientist's primary tools, does not work in it." (1990, p. 2)

The domain of ethical reasoning can be viewed as a *weak analytic domain* characterized by the following attributes. First, the given "rules" (i.e., laws, codes, or principles) are available almost exclusively at a highly conceptual, abstract level. This means that the rules may contain *open-textured* terms (Twining and Miers, 1976; Gardner, 1987). That is, conditions, premises, or clauses that are not precise or that cover a wide range of specific facts, or are highly subject to interpretation and may even have different meanings in different contexts. A second characteristic of weak analytic domains, closely related to the first characteristic, is that the actions prescribed by the given rules, i.e., the rules' conclusions, may also be abstract. Thus, even if one is able to determine that a particular rule applies to a given fact situation, the rule's consequent recommendation may be difficult to execute because it is highly conceptual or vague. For instance, how does one determine the action prescribed by NSPE code I.1., a principle used by SIROCCO to perform its reasoning, in which professional engineers are urged to "hold paramount" the safety, health, and welfare of the public? The prescribed action is clearly tied to the specific circumstances of a case to which it is applied. Third, abstract rules often conflict with one another in particular situations with no deductive or formal means of arbitrating such conflicts. That is, more than one rule may appear to apply to a given fact situation, but neither the abstract rules nor the general knowledge of the domain provide clear resolution.

Another important lesson from the Truth-Teller and SIROCCO projects is the sheer difficulty in imbuing a computer program with the sort of flexible intelligence required to perform ethical analysis. While both programs performed reasonably well in the studies mentioned above, neither could be said to have performed at the level of an expert human at the same task. While the goal was not to emulate human ability (or take the task of ethical analysis away from the human), it is important for computational artifacts that purport to support ethical reasoning to at least perform well enough to encourage humans to use the programs as aids in their own reasoning.

It is important to make clear that the author's belief that computer programs can only act as aids in ethical reasoning is not due to a high regard for human ethical decision making. Of course, humans often make errors in ethical reasoning. Rather, the author's position is based, first of all, on the existence of so many plausible, competing approaches to ethical problem solving (e.g., utilitarianism (Mill, 1979), respect for persons ethics (Kant, 1959), reflective equilibrium (Goodman, 1955; Rawls, 1971)). Which philosophical method can be claimed to be the "correct" approach to ethical reasoning in the same sense that calculus is accepted as a means of solving engineering problems or first-order logic is used to solve syllogisms? It is difficult to imagine that a single ethical reasoning approach embodied in a single computer program could deliver a definitive approach to ethical reasoning. Second, it is presumptuous to think that the subtleties of any of the well-known systems of ethics could be fully implemented in a computer program. Finally, there is an ethical dimension to the author's view. Is the human race ready to fully relegate human ethical decision making to machines? This seems highly doubtful.

Future Directions

The author's most recent work and interest has been in the area of intelligent tutoring systems (McLaren *et al*, 2005; Koedinger *et al*, 2004). As such, the author has started to investigate whether case comparisons, such as those produced by Truth-Teller, could be used as the basis for an intelligent tutor. The idea is to explore whether Truth-Teller's comparison rules and procedures can:

- be improved and extended to cover the kinds of reasons involved in comparing more technically complex cases, such as those tackled by SIROCCO, and
- serve as the basis of a Cognitive Tutor to help a student understand and perform the phases taken by the Truth-Teller program.

Cognitive Tutors are based on Anderson's ACT-R theory (Anderson, 1993), according to which humans use production rules, modular IF-THEN constructs, to perform problem-solving steps in a wide variety of domains. Key concepts underlying Cognitive Tutors are "learn by doing," helping a student learn by engaging her in actual problem solving, and immediate feedback, providing guidance to a student at the time they request a hint or make a mistake. For domains like algebra, the production rules in a cognitive model indicate correct problem-solving steps a student might take but also plausible incorrect steps. The model provides feedback in the form of error messages, when the student takes a step anticipated by a "buggy rule," and hints, when the student asks for help.

Developing a Cognitive Tutor for case comparison presents some stiff challenges, not the least of which is that, unlike previous domains in which Cognitive Tutors have been used, such as algebra and programming, in practical ethics answers are not always and easily identified as correct or incorrect, and the rules, as explained earlier, are more abstract and ill-defined. As a result, while learn by doing fits ethics case comparison very well, the concept of immediate feedback needs to be adapted.

Much more than the rules of algebra, the "rules" of ethics case comparison are more abstract descriptions of a process. If followed, they can help a student frame an intelligent comparison. If not followed, it is not necessarily an indication of failure. Unlike algebra, answers may be nuanced rather than simply right or wrong, and the Cognitive Tutor approach must be adapted accordingly to help students frame comparisons and identify and compare reasons.

The main point is that the production rules employed in Truth-Teller's first three phases, particularly the Qualification phase, provide a core set of rules that can be improved and recast as a set of rules for comparing cases within a Cognitive Tutor framework. A planned empirical study of case comparisons, involving more technically complex ethics cases, will enable refinement and augmentation of these comparison rules. At the same time, the empirical study of subjects' comparing cases may reveal plausible misconceptions about the comparison process that can serve as buggy rules, or faulty production rules that present opportunities to correct the student.

A related direction is exploring whether the priority rules of Ross' theory of *prima facie* duties (1930), such as non-maleficence normally overriding other duties and fidelity normally overriding beneficence, might benefit the Truth-Teller comparison method. At the very least it would ground Truth-Teller's approach in a more established theory (currently priority rules are implemented in a somewhat *ad hoc* manner, based loosely on Bok (1989)). Such an extension to Truth-Teller would also benefit the planned Cognitive Tutor, as explanations to students could be supported with reference to Ross's theory.

Acknowledgements. Kevin Ashley contributed greatly to the ideas behind both Truth-Teller and SIROCCO. This work was supported, in part, by NSF-LIS grant No. 9720341.

References

Anderson, J. R. (1993). *Rules of the Mind*. Mahwah, NJ: Lawrence Erlbaum.

Aristotle, (edited and published in 1924) *Nicomachean Ethics*. W. D. Ross, editor, Oxford, 1924.

Arras, J. D. (1991). Getting Down to Cases: The Revival of Casuistry in Bioethics. In *Journal of Medicine and Philosophy*, 16, 29-51.

Ashley, K. D. (1990). *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*. Cambridge: MIT Press, 1990.

Ashley, K. D. and McLaren, B. M. (1994). A CBR Knowledge Representation for Practical Ethics. In the *Proceedings of the Second European Workshop on Case-Based Reasoning*, *(EWCBR)*, Chantilly, France, 1994.

Ashley, K. D. and McLaren, B. M. (1995). Reasoning with Reasons in Case-Based Comparisons. In the *Proceedings of the First International Conference on Case-Based Reasoning*, Sesimbra, Portugal.

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

Brody, B. (2003). *Taking Issue: Pluralism and Casuistry in Bioethics*. Georgetown University Press.

Gardner, A. (1987). *An Artificial Intelligence Approach to Legal Reasoning*. Cambridge, MA: MIT Press.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Harris, C. E., Pritchard, M. S., and Rabins, M. J. (2004). *Engineering Ethics: Concepts and Cases*. 3rd Edition. Wadsworth, a division of Thomson Learning.

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

Kant, I. (1959). *Foundations of the Metaphysics of Morals*. New York: Liberal Arts Press, Bobbs-Merrill.

Koedinger, K., Aleven, V., Heffernan, N., McLaren, B. M., and Hockenberry, M. (2004). Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration; In the *Proceedings of the Seventh International Conference on Intelligent Tutoring Systems* (ITS-2004).

Lewis, D. D., Schapire, R. E., Callan, J. P., and Papka, R. (1996). Training Algorithms for Linear Text Classifiers. In *Proceedings of the 19th Ann. Int'l ACM-SIGIR Conference on Research and Development in Information Retrieval*. Zurich.

McLaren, B. M. and Ashley, K. D. (1995a). Case-Based Comparative Evaluation in Truth-Teller. In the *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Pittsburgh, PA.

McLaren, B. M. and Ashley, K. D. (1995b). Context Sensitive Case Comparisons in Practical Ethics: Reasoning about Reasons. In the *Proceedings of the Fifth*

International Conference on Artificial Intelligence and Law, College Park, MD.

McLaren, B. M. and Ashley, K. D. (2000). Assessing Relevance with Extensionally Defined Principles and Cases; In the *Proceedings of AAAI-2000*, Austin, Texas.

McLaren, B. M. (2003). Extensionally Defining Principles and Cases in Ethics: an AI Model; *Artificial Intelligence Journal*, Volume 150, November 2003, pp. 145-181.

McLaren, B. M., Bollen, L., Walker, E., Harrer, A., and Sewall, J. (2005). Cognitive Tutoring of Collaboration: Developmental and Empirical Steps Toward Realization;, In the *Proceedings of the Conference on Computer Supported Collaborative Learning* (CSCL-05), Taipei, Taiwan in May/June 2005.

Mill, J. S. (1979). *Utilitarianism*. G. Sher, Ed. Indianapolis, IN: Hackett.

Rawls, J. (1971). *A Theory of Justice*, 2nd Edition 1999, Cambridge, MA: Harvard University Press.

Ross, W. D. (1930). *The Right and the Good*. New York: Oxford University Press.

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

Toulmin,S. E. (1958). *The Uses of Argument*. Cambridge, England: Cambridge University Press.

Twining, W. and Miers, D. (1976). *How to Do Things With Rules*. London: Cox and Wyman, Ltd.