Title

Computing Value Judgements During Story Understanding

Permalink

https://escholarship.org/uc/item/3nv5v965

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 11(0)

Author

Reeves, John F.

Publication Date

1989

Peer reviewed

Computing Value Judgments During Story Understanding

John F. Reeves

Computer Science Department
University of California, Los Angeles

ABSTRACT

During story understanding readers make value judgments—judgments of the 'goodness' or 'badness' of characters' actions. This paper presents the representational structures and processes used to make value judgments by the computer program THUNDER. THUNDER creates evaluative beliefs about characters' plans based on a set of universal pragmatic and ethical judgment rules. To account for subjective differences in evaluative belief, THUNDER has a specific ideology to represent the idiosyncratic aspects of evaluation. There are two components in the representation of ideology: (1) a set of important, long term goals called values, and (2) a collection of planning strategies for each value. This representation for ideology allows THUNDER to reason about what is 'good', and what it believes to be 'good ways to get what is good.' The representation and rules for value judgments are used to (1) make inferences about character belief and ideology, (2) represent expectation knowledge based on personality traits, and (3) reason about the obligations that characters acquire.

INTRODUCTION

Plan evaluation is the process of deciding whether or not a plan should be used. Two types of reasons are used in plan evaluation: (1) *pragmatic* reasons, reasons about the consequences of the plan for the planner, and (2) *ethical* reasons, reasons about the consequences of the plan for people other than the planner. As an example of the two types of reasons, consider the reasons that the following two plans are 'bad':

EX-1: To save money, John decided never to change the oil in his new car.

EX-2: To get the money to buy a new car, John decided to rob a bank.

In EX-1, John's plan to save money is 'bad' because it will end up costing him more to replace the car engine when the bearings seize up than to perform regular maintenance. In EX-2, John's plan to get money is 'bad' because of the loss of property he is causing to the bank depositors. These two senses of the word "bad" correspond to the questions (1) will the plan work? and (2) is the plan ethically right?

THUNDER (Thematic UNDerstanding From Ethical Reasoning) (Reeves, 1988) is a story understanding program that reads short narratives and answers ethical and thematic questions. Value judgment is the primary task of THUNDER during story understanding. THUNDER uses its value judgments to recognize *belief conflict patterns*: abstract situations where the ethical judgments of the reader and story characters conflict. THUNDER uses belief conflict patterns to (1) organize the representation of the story, (2) focus attention on the thematically interesting elements of the story, and (3) identify the theme of the story by resolving the belief conflict. By judging story character's plans, THUNDER can answer the following questions (from EX-1 and EX-2, respectively):

> Why was John wrong not to put oil in his car?

It is wrong because not putting oil in the car will damage the car, and the car is more expensive than the oil.

> Why was John wrong to rob the bank?

It is wrong because John is taking money from the bank depositors.

In order to make evaluative judgments about story characters' actions, THUNDER has to (1) have knowledge about what is 'good' and 'bad', and (2) be able to reason about how actions are evaluated. THUNDER's *evaluative beliefs* about different types of human goals and ways for achieving those goals represent knowledge about normative value. Evaluative beliefs are organized in THUNDER's *ideology*. Expected goal successes and failures are used to determine normative goodness; goal successes are evaluated positively, and goal failures are evaluated negatively. To make evaluations of situations, THUNDER has a set of *judgment rules* which are applied to situations to create evaluative beliefs about story characters and what they have done.

This paper presents the structures and processes that are used in THUNDER's evaluative reasoning model. When references are made to the program's beliefs or values, these terms refer to the data structures that are used in the computer program to implement psychological functions. The usefulness of these structures is shown by (1) using them to make value judgments, and (2) identifying the components of the value judgment process. In addition to making judgments, the processes in the model are used to represent and reason about personality traits and obligation.

FACTUAL AND EVALUATIVE BELIEF

Representing and reasoning about beliefs is a fundamental problem for Artificial Intelligence (AI) systems. Previous approaches have addressed belief in terms of uncertainty and truth maintenance (e.g. (Cohen, 1985; Pearl, 1988)). However, there is more to belief than just the degree of certainty with which a proposition is held. Part of the problem is that these systems have not made the distinction between *factual* and *evaluative* belief. Factual beliefs are evaluated in terms of truth, and evaluative beliefs are evaluated in terms of 'good' and 'bad.'

In THUNDER, a *belief* is a conceptual object attributed to a person. The person (the believer) can be the reader/system or a story character. The *content* of a belief is a constituent conceptual object that the belief is about, such as a plan or a future state of affairs. The *strength* of a belief is the degree on either the factual or evaluative scale with which the believer holds the belief.

A *factual belief* is a belief that the content of the belief is true or false (has a truth strength). Factual beliefs can be held with degrees of certainty or probabilities.

An *evaluative belief* is a belief that the content of the belief is positively or negatively evaluated (has a positive or negative strength). Most of the evaluative beliefs used in THUNDER are about plans; evaluative beliefs about actions are handled by reference to beliefs about the plans in which they are a part. Positive and negative evaluations of plans correspond to beliefs that the plan should or should not be used, respectively.

A *judgment* is an evaluative belief that a person creates and is the product of a judgment process. A pragmatic judgment is the creation of a evaluative belief for pragmatic reasons, and an ethical judgment is the creation of an evaluative belief for ethical reasons. The process of making an ethical judgment is termed *ethical reasoning.*

The distinction between factual and evaluative beliefs is a metaethical philosophic position called *noncognitivism* (Boyce and Jensen, 1978, pp. 76-81). The basic precepts of noncognitivism are:

- Evaluative statements are not evaluated in terms of truth.
- There is no method of ultimate justification of evaluative statements (as in scientific or mathematical proof).
- The function of evaluative statements is to express emotions (Ayer, 1935), to influence other's attitudes (Stevenson, 1944), or to rationally guide human conduct (Hare, 1952).

The problem for noncognitivist philosophers is defining how evaluative statements are justified, and what constitutes a good reason for holding a evaluative belief (Toulmin, 1950). In the construction of THUNDER, 'good' is defined in terms of the values of an individual, and then character actions are evaluated in terms of the values. Using different value systems will produce different evaluations. For example, a Catholic and a Samurai would have different evaluations of the following story:

EX-3: A high school student killed himself after flunking out of school.

The Catholic believes that suicide is a mortal sin, and therefore the student's plan is wrong, but the Samurai believes that death is preferable to living in disgrace.

Once THUNDER makes an evaluative judgment, it is not concerned with establishing the truth of the statement, but rather with the reasons for the judgment, and how the judgment can be used in story understanding.

MODELING READER IDEOLOGY

An *ideology* is an organization for goals and plans in memory based on evaluative beliefs about states that should be desired, and how to go about achieving those states. The representation for ideology in THUNDER has two components: (1) the *value system*, a set of abstract, high-level evaluative beliefs about goals (called *values*) ordered by their relative importance (Rokeach, 1973), and (2) a set of *planning strategies* for each value, representing the ways that a person believes the value should be achieved. The values are based on Rokeach's terminal human values (Rokeach, 1973), and represented in terms of Schank and Abelson's goal primitives [1977].[1]

[1] In the notation used for goals, the goal type is signified by the letter preceding the goal name. Achievement goals (A) are a motivations to attain valued acquisitions

THUNDER's representation for ideology is an extension of Carbonell's system in the POLITICS program(Carbonell, 1978; Carbonell, 1979) where ideologies were represented by *goal trees*. A goal tree is a hierarchy of goals ordered by *subgoal* and *relative importance* relations. THUNDER divides the goal importance and instrumentality functions of ideology into separate structures: the value list and planning strategies, respectively. By making this separation, THUNDER loses the advantage of having one unified structure for representing ideology (the goal tree), but is able to reason more effectively about the end states that the program believes are 'good' and about the value of types of plans.

Another difference between POLITICS and THUNDER is that THUNDER makes a distinction between the role of pragmatic and ethical belief in the representation of ideology. POLITICS evaluated the consequences of events in terms of goal trees, so a 'good' plan was an effective plan for an important goal which avoids failures of other important goals. This is only the pragmatic side of evaluative belief—an ethical plan evaluator also has to consider the goals and goal failure effects on parties other than the planner. In THUNDER, the value system represents only the relative importance of the reader's *values*, and general judgment rules are used to evaluate goal successes and failures. Thus, the value system doesn't include instrumental goals, and separates the concept of 'what is good' from 'good ways to get what is good.'

Values

There are two types of values: (1) *preservation* values, and (2) *achievement* values. Preservation values are the things that everyone wants to keep: their health, freedom, possessions, self esteem, and social esteem. Achievement values are the things that people want out of life, such as a successful career, spiritual salvation, or excitement and good times. Preservation values are things that people should not have threatened, or worse, have fail, while achievement values are the things that people believe are valuable to try to get.

Preservation values are positive evaluative beliefs about having and keeping something. Having a preservation value allows the individual to evaluate actions that threaten the value. For example, if a person values their health, then threats and damage to their health are evaluated negatively. In a value system, preservation values can be held for a particular group of people: the individual, their family, friends, a social group (a community, nation, or race), or everybody. A white supremacist, for example, believes that freedom should be preserved only for the Caucasian race, and a patriot believes in freedom preservation for his nation.

Achievement values are positive evaluative beliefs about things that people try to get. There are four classes of achievement values: (1) *acquirement* values, like achieving power, money, or status, (2) *personal* values, such as achieving salvation, tranquillity, or wisdom, (3) *interpersonal* values, such as achieving love, respect, or friendship, and (4) *entertainment* values, such as pleasure, excitement, or enjoyment.

THUNDER's value system contains preservation and achievement values ordered by their relative importance. There are five preservation values: P-Health, P-Freedom, P-Possessions, P-Self-esteem, P-Social-esteem. Each preservation value is believed to be important for an ordered list of people: THUNDER, family, friends, social group, nation, and everyone else. The achievement values are less important than the preservation values, and are in the following order: the interpersonal values A-Love and A-Friendship, the personal values A-Intellect, A-Tranquility, and A-Salvation, the acquirement values A-Respect, A-Status, A-Power, and A-Possessions, and the entertainment values E-Excitement, and E-Enjoyment. Note that the value system can be reordered to represent different ideologies. For example, a spiritual person would have A-Salvation relatively higher in the list, while a patriot might have P-Freedom(nation) above the other preservation values.

Because there are a small set of goal primitives that are used in the value system, the system can monitor the goals for activation, threats, and failures. There are three points to notice in the construction of the value system: (1) the set of values that the program has is fairly short (Rokeach, 1973), (2) not all goals that the program knows about are included in the value system, and (3) the value system does not represent instrumental relations between the values because the system represents what is valuable, not how to maintain or achieve those valuable states.

Since the value system represents what the reader believes to be valuable, actions that threaten or cause values to fail are believed to be bad. When a story character does something that threatens a goal or causes a goal failure, the character's plan can be evaluated for how bad the plan is. In the evaluation, the following factors are used:

1. *Importance of the failed goal.* How important is the failed goal to the person whose goal it is? It is worse to violate an important goal of a person than a less important goal.
2. *Duration of the goal failure.* How bad is the

or social positions. Other goal types are preservation (P), delta (D), and enjoyment (E).

goal failure? The duration of a goal failure can be measured by how hard it is to recover: loss of property is replaceable, but consumes resources that the person suffering the goal failure would not have had to expend. Some goal failures are non-recoverable, such as loss of life.

3. *Scope of the goal failure.* How many goal failures does the plan cause? If John punches Jerry, he has caused a goal failure for just one person, but if John dumps toxic waste in the old swimmin' hole, he has causes goal failures for anyone who wants to use the swimmin' hole in the future.

Values are positive evaluative beliefs about goals. When THUNDER makes judgments about goal failures and successes it uses the goals that are in the values in the value system. Successes and failures of the goals in the value system are called *value successes* and *value failures*, respectively.

Planning Strategies

The second element of ideology involves beliefs about the ways that the values should be achieved. For example, in EX-1 John believes that a good way to save money is to not put oil in his car. John's plan is an instance the general strategy of being thrifty: John believes that a good way of possessing money is to avoid spending it. From EX-1, the reader knows that John values saving money, and also how he goes about saving money. An alternate strategy would be to avoid risking the money, as in:

> EX–4: To save money, John invested in treasury bonds.

Planning strategies are associated with the various values in the ideology to make the distinction between value and instrumentality. Planning strategies are evaluative beliefs about plans for values in the value system. The content of a planning strategy is an abstract kind of plan, such as plans involving prevention for preservation of health.

By associating planning strategies with values, the system can quickly find good plans for a given value. Planning strategies can be used to organize plans for values by providing intermediate nodes in a plan hierarchy where plans are indexed by the ways in which the planner believes that they are valuable. If, for example, a person believes that prevention is a positive pragmatic strategy for preservation of health, then specific plans for preservation of health can be indexed under the planning strategy by their appropriate context, such as going to the doctor regularly, exercising and eating right, and avoiding situations were their health can be threatened. By organizing plans by the values that they achieve, and by their relative value, the system can reason about instrumental relations between the plans and planning trade-offs. For example, a person who believes in the prevention-for-health strategy will not believe that health threatening activities (such as skydiving or hangliding) are effective plans for entertainment, and that a good doctor is worth an additional cost.

PRAGMATIC AND ETHICAL REASONING

Evaluation of a person's actions consists of creating evaluative beliefs about their plans. In order to create evaluative beliefs, THUNDER has to (1) understand what the character is doing (his plan) and why he is doing it (his goal), (2) evaluate the character's plan by generating reasons for an evaluative belief about the plan, and (3) generate the reasons for the character's positive evaluative belief about the plan.

In EX-2, there are three pragmatic reasons that John's plan for getting money by robbing a bank is positively evaluated: (1) it helps achieve his goal of buying a car by getting a lot of money, (2) the plan has low resource consumption—it takes less time than working for the money, and is less expensive than investing, and (3) the plan is highly effective—better than mugging or robbing a 7-11. There is one pragmatic reason that robbing a bank is negatively evaluated: the liability of capture and imprisonment. Ethically, robbing a bank is wrong because of (1) the loss suffered by the people who have their money in the bank, and (2) the threatened loss of life to the people who were working in the bank.

Each reason for an evaluation of a plan can be broken down into two components: (1) a factual belief about the plan, and (2) a pragmatic or ethical *judgment rule* that is used to derive an evaluative belief from a factual belief. To generate appropriate factual beliefs about a character's plan, the following factors have to be considered:

1. *Plan availability.* What other plans are available to the planner? What are the relative merits of the other available plans?
2. *Goal importance.* How important is the planner's goal? If the plan causes goal failures for others, how important are the goals that fail?
3. *Intention.* If the plan causes goal failures, does the character realize that he is causing a goal failure? If a character is executing an action that will cause a goal failure for himself, such as locking the car door with the keys inside, then the action should be evaluated as stupid, but not as evil.

When THUNDER reads about a character's action, it infers a plan that that action is a part of.

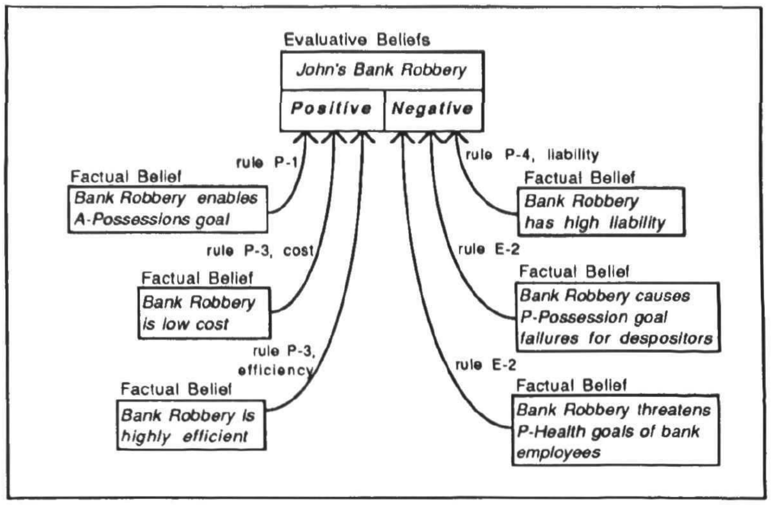

Figure 1: Pragmatic and Ethical Reasons for THUNDER's Evaluative Belief about Bank Robbery

Once a belief graph has been constructed, THUNDER makes a determination of its actual evaluative belief about the plan. In the determination, ethical reasons take precedence over pragmatic reasons. Because there is an ethical reason that John's bank robbery is evaluated negatively, THUNDER holds the evaluative belief that the bank robbery is wrong, even though there are pragmatic reasons that bank robbery is positively evaluated.

INFERRING CHARACTER BELIEFS AND IDEOLOGY

In addition to creating its own evaluative belief about the character's plan, THUNDER has to figure out why the character believed that the plan was justified. For a plan that THUNDER believes is pragmatically wrong, THUNDER can make inferences about the character's beliefs. For a plan that THUNDER believes is ethically wrong, inferences about the character's ideology can be made. When a character executes a plan that is evaluated negatively for pragmatic reasons, THUNDER uses the following *pragmatic inference rules* (PI rules):

> PI-1: The character doesn't have the factual belief about the plan that THUNDER used to make its evaluative assessment.
>
> or
>
> PI-2: The character believes that the goal that they are achieving is more important than the goal that they are causing to fail.

These two inference rules are mutually exclusive, and depend on the character's intention. For example, in EX-1 either John doesn't know that not changing the oil in his car will damages the engine (rule PI-1), or he knows it and believes that the short term goal success of saving money by not changing the oil is more important than the long term goal failure of having to buy a new car (rule PI-2).

When a character executes an ethically wrong plan, the following *ethical inference rules* (EI rules) are used:

> EI-1: The character believes that their value is more important than the value failure that they caused.
>
> and
>
> EI-2: The character believes that the ethically wrong plan is the only way to achieve their value.
>
> or
>
> EI-2': The character believes that the ethically wrong plan is a less expensive (in time or resources) way of achieving their value than other available plans.

These inferences are based on observations that people do not go out of their way to do ethically wrong actions; they must have a motivation (rule EI-1) and a rationale (rules E1-2 and EI-2'). From these inferences about what the character believes to be valuable, THUNDER can begin to construct the character's ideology. From EX-2, THUNDER can infer that John believes that his goal of getting a car is more important than the bank depositor's goal of preserving their money, and that John has pragmatic beliefs about bank robbery that make it better than other available plans.

If the goal of the plan is not an instance of a value in THUNDER's value system, THUNDER assumes that the goal is the subgoal of a larger plan schema, and continues to infer plans until a plan for a value is found. For example, if THUNDER reads:

EX-5: John robbed a bank.

THUNDER infers that that John is robbing the bank to get money. However, the goal of getting money is not a value. So THUNDER continues to find plans that are instrumentally enabled by getting money, such as A-Possessions plans, or plans that bank robbers use their money to pursue, such as plans for entertainment goals (by spending the money on parties or drugs.)

The inferred plan chain from action to a value is made up of one or more individual plan schemas. Each plan schema contains the goal failures that the plan causes; for example the 'threaten' plan schema (PS-Threaten) contains the motivated P-Health goal for the person threatened. PS-Threaten is a sub-plan of the bank robbery schema (PS-Bank-robbery), so when John robs a bank, THUNDER knows that the bank employees are suffering a P-Health value failure. The goal at the top of of the complete plan is called the *value of the plan*, or the value that the plan achieves. When complete plans are inferred, they can be evaluated both for planning failures (Dyer, 1983) and the ethical and pragmatic consequences. Even thought the plan may not have been completely executed, an evaluation can be made from the values that reader expects to succeed and fail.

THUNDER uses the following pragmatic judgment rules for evaluating story characters' plans:

> P-1: If plan P1 achieves its value, then P1 is positively evaluated.
>
> P-2: If plan P1 causes value failure VF for the planner, then P1 is negatively evaluated.
>
> P-3: If plan P1 is better on plan metric[2] I than competing plan P2, then P1 is positively evaluated.
>
> P-4: If plan P1 is worse on plan metric I than competing plan P2, then P1 is negatively evaluated.

The following ethical judgment rules are used in making ethical judgments about plans:

> E-1: If plan P1 achieves value V for another party, then P1 is positively evaluated.
>
> E-2: If plan P1 causes value failure VF for another party, then P1 is negatively evaluated.
>
> E-3: If plan P1 achieves value V while intentionally causing value failure VF and V is more important than VF, then P1 is positively evaluated.
>
> E-4: If plan P1 achieves value V while intentionally causing goal failure VF and V is less important than VF, then P1 is negatively evaluated.

The reasons that rules E-3 and E-4 are ethical, rather than pragmatic, is that even if both of the goals in V and VF are the planner's goals, the importance measure is the understander's. For example:

EX-6: John took steroids to improve his physique.

If John is understood to be improving his physique to feel better about himself, and both John and reader knows about the harmful side effects of steroid usage, then V is John's P-Self-esteem goal, and VF is John's P-Health goal. Rules P-1 and P-2 evaluate the pragmatic consequences of the plan, while E-4 evaluates the plan as unethical because THUNDER believes that John should value his health more than his self-esteem.

These judgment rules can serve as deductive rules in plan evaluation, and as preferences or advice in plan selection or creation. For example, rule E-2 says to prefer plans that do not cause goal failures for others over those that do. Using these rules and factual beliefs about bank robbing, a belief graph can be constructed for EX-2 as shown in figure 1.

The beliefs in figure 1 are THUNDER's. The factual beliefs are THUNDER's knowledge about bank robbery, and how bank robbery can be compared to other plans for getting money. The links between factual and evaluative beliefs are labeled by judgment rules. For example, one reason that THUNDER has for believing that John's bank robbery is positively evaluated is that (1) THUNDER believes that the bank robbery will help achieve the A-Possessions goal by providing John with the money to buy a car (the factual belief), and (2) there is a pragmatic rule that plans that achieve values are positively evaluated (judgment rule P-1). Notice that THUNDER has reasons for both a positive and negative evaluation of John's bank robbery. However, THUNDER is not holding contradicting beliefs, but has reasons for believing both sides of the evaluation.

[2] A *plan metric* (Dyer, 1983) is a measurement unit for plans. For example, the "cost" metric measures how many resources are used during plan executions. Other plan metrics are enablement, efficacy, risk, coordination, availability, legitimacy, affect, skill, vulnerability, and liability.

VALUE JUDGMENTS ABOUT CHARACTERS

Value judgments about story characters can be represented by value judgments about the plans that the characters execute or will be expected to execute. Carbonell [1980] recognized the relationship of values to personality traits. In his model of personality traits, Carbonell used a prototypical goal tree to represent the normative orientation of people's goals. Personality traits were then represented as modifications to the prototypical goal hierarchy. For example, the modifications to the goal tree for an "ambitious" person are to have their achievement goals moved higher in the tree, and preservation goals for others moved lower. This represents that an ambitious person will sacrifice family and friends to get ahead. Carbonell [1980, p.67] notes that goal trees do not completely represent personality traits; some traits have *means-oriented* components, meaning that they describe the planning choices that a character is expected to make. An "ambitious" person is expected to use deceptive plans, and will be hesitant to compromise, while a "capable" person will make correct decisions in plan selection and carry out plans without making errors.

The means-oriented components of personality traits can be represented by including the method by which a character achieves goals, or causes goals to fail. In THUNDER, the reasons that a character is expected to do 'good' or 'bad' actions are represented by *character assessments*. Character assessments are representations of the reasons for evaluative beliefs about characters, and provide the reader with a moral context in which to judge their actions.

There are two types of character assessments, corresponding to the ends of the evaluative scale: (1) *positive* character assessments, that represent how the character achieves goal successes, and (2) *negative*, that represent how the character causes goal failures. Character assessments have three components: (1) the type of goal that the character will achieve, or cause to fail, (2) the planning situation in which the assessment applies, and (3) the action that the character does in that situation to cause the goal consequences. For example, in the negative assessment for a "coward", the goal that the person will have fail is preservation of self esteem, the plan-situation where the failure occurs is during plan-execution in reaction to adversity, and the method of failure is that person abandons their goal when faced with an adverse situation. In contrast, an "imaginative" person has a positive assessment for all goals that apply in plan creation situations, and an "affectionate" person has a positive assessment for achieving other people's friendship and love goals by executing plans for those goals.

There are two sources of character assessments in story understanding: (1) *direct* character assessments, which are generated from the goal successes and failures that characters have in the story, and (2) *background* character assessments, which are associated with lexical entries, such as "coward", "affectionate, and "imaginative", or with other knowledge sources containing expectations about people, such as Schank and Abelson's [1977] role themes.

Direct character assessments provide reasons for evaluative beliefs about characters from goal successes and failures in the story, and background character assessments provide reasons based on a character's *capability* to cause goals to succeed or fail. For example, compare:

EX–7A: John beat up Jerry and took his lunch money.

EX–7B: John was a mean, spiteful sixth grader.

In EX–7A, John is bad because of what he did: a direct negative character assessment is built because he violated Jerry's P-HEALTH goal. In EX–7B, John is bad because the reader expects him to do things like beat people up; based on his description as mean and spiteful, a background negative character assessment is built for John that represents the expectation that John will cause P-HEALTH goal failures for others.

Character assessments provide (1) reasons for the reader's evaluation of the character, and (2) expectations about future character behavior. The expectation information associated with personality traits can be accessed from static knowledge of background assessments, or created dynamically by THUNDER as direct character assessments. Thus, expectations can be generated both from character descriptions and their actions.

REASONING ABOUT OBLIGATION

In addition to having goals, story characters may incur *obligations*. An obligation is a belief that someone should have a goal, but not that that person necessarily has that goal. An obligation is represented as a positive evaluative belief where the content of the belief is that the character has a goal. For example, if THUNDER reads the sentence:

EX–8A: John borrowed $5 from Bill...

From knowledge about 'borrowing', THUNDER knows that John has an obligation to pay Bill back. This obligation is represented as a goal that John should have, so THUNDER positively evaluates the situation where John has the goal of paying Bill $5. John may not share the belief; if the sentence continued:

EX–8B: ..., which John never intended to pay back.

THUNDER would make the judgment that John's intention not to repay the loan is ethically wrong, because THUNDER has the belief that John should have the goal, but John does not have the goal.

Since the content of an obligation is a goal for another party, characters that achieve these goals are evaluated positively for ethical reasons. Similarly, characters that violate obligations are evaluated negatively for ethical reasons. Story characters acquire obligations from the relationships that they become involved in, and from their description. Obligations are associated with knowledge structures for relationships, such as 'lovers', 'teacher/student' and 'employer/employee', and role-themes, like 'policeman' or 'bank president.' For example, in the 'teacher/student' relationship, the teacher has the goal that the students learn the material, and the students have the goal of showing the teacher that they have learned the material. Thus, one ethical reason that cheating on a test is wrong is that it violates the student's obligation to the teacher.

The values and obligations that THUNDER believes are good are distinguished from the goals that characters have, so that what a character wants and what the system believes that a character *should* want do not get confused. The goals that a character has provide their motivations, and the THUNDER's values and understanding of obligations provide a moral context in which to evaluate the character's actions.

CONCLUSIONS

The process of making value judgments has been implemented in THUNDER by modeling the creation of evaluative beliefs. Story characters' plans are evaluated using a general set of pragmatic and ethical judgment rules. These rules are independent of any particular individual. The parts of the model that are idiosyncratic to the individual are the data that the rules operate on: the factual and evaluative beliefs that the system has. THUNDER's primary task during story understanding is to make evaluative judgments about story characters' actions, and then to use those judgments to (1) focus attention, (2) control inferencing, and (3) recognize the thematic elements of the story. In addition, the representation for the reasons for evaluation of characters' plans can also be used to represent (1) the reasons for evaluation of characters, and (2) expectations and rationale for character behavior. By representing obligations as beliefs about goals for others, THUNDER can reason about violations of interpersonal relations.

ACKNOWLEDGMENTS

This work is supported in part by a grant from the Hughes Artificial Intelligence Center. Thanks to Dr. Michael Dyer, Jack Hodges, Trent Lange, Seth Goldman, Stephanie August, and the reviewers for their comments on an earlier version of this paper.

REFERENCES

Ayer, A. J. (1935). *Language, Truth and Logic.* Dover Publications, New York, second edition.

Boyce, W. D. and Jensen, L. C. (1978). *Moral Reasoning: A Psychological-Philosophical Integration.* University of Nebraska Press, Lincoln, NB.

Carbonell, J. G. (1978). Politics: Automated ideological reasoning. *Cognitive Science*, 2(1):29–51.

Carbonell, J. G. (1979). *Subjective Understanding: Computer Models of Belief Systems,.* PhD thesis, Department of Computer Science, Yale University, New Haven CT. Technical Report 150.

Carbonell, J. G. (1980). Towards a process model of human personality traits. *Artificial Intelligence*, 15:49–74.

Cohen, P. R. (1985). *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach (Research Notes on Artificial Intelligence 2).* Pitman Advanced Publishing, London.

Dyer, M. G. (1983). *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension.* MIT Press, Cambridge, MA.

Hare, R. M. (1952). *The Language of Morals.* Oxford University Press, Oxford.

Pearl, J. (1988). *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference.* Morgan-Kaufman, San Mateo, CA.

Reeves, J. F. (1988). Ethical understanding: Recognizing and using belief conflict in narrative understanding. In *Proceedings of AAAI-88*, St Paul, MN.

Rokeach, M. (1973). *The Nature of Human Values.* Free Press, New York.

Schank, R. C. and Abelson, R. P. (1977). *Scripts Plans Goals and Understanding.* Lawrence Erlbaum, Hillsdale, NJ.

Stevenson, C. L. (1944). *Ethics and Language.* Yale University Press, New Haven, CT.

Toulmin, S. (1950). *The Place of Reason in Ethics.* Cambridge University Press, Cambridge.