Proceedings, The Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment

Representing Morals in Terms of Emotion

Margaret Sarlej and **Malcolm Ryan**
School of Computer Science and Engineering
University of New South Wales
Sydney, NSW, Australia
msarlej@cse.unsw.edu.au, malcolmr@cse.unsw.edu.au

Abstract

Morals are an important part of many stories, and central to why storytelling developed in the first place as a means of communication. They have the potential to provide a framework for developing story structure, which could be utilised by modern storytelling systems. To achieve this we need a general representation for morals. We propose patterns of character emotion as a suitable foundation. In this paper, we categorise Aesop's fables based on the morals they convey, and use them as a source of emotion data corresponding to those morals. We use inductive logic programming to identify relationships between particular patterns of emotion and the morals of the stories in which they arise.

Introduction

Storytelling systems can be divided into two main categories: character-centric and author-centric. Character-centric systems attempt to simulate autonomous character agents which interact to produce a story. For example, *Virtual Storyteller* (Theune et al. 2003) develops plot using semi-autonomous agents whose behaviour is guided by their emotional state. Author-centric systems, on the other hand, focus on an author's goals in writing a story. These can include structural goals or intended effects on readers. For example, Cheong and Young's *Suspenser* selects which parts of a story to tell, and in what order, to create feelings of suspense (Cheong and Young 2008). The primary goal of both approaches is the same: to generate stories that are coherent and interesting, mainly targeted at entertainment.

However, it should be noted that historically the purpose of storytelling was not just to amuse and entertain, but to convey a lesson or point: the moral of the story (Ryan 1991). Consequently, stories were often designed around the particular moral or message they were intended to communicate. It is important that this aspect of storytelling is not neglected by storytelling systems, but leveraged as a possible framework for developing story structure; morals are essentially an author-centric device. The goal of this work is to develop an approach for representing morals so they can be incorporated as a feature of storytelling systems.

Morals in Storytelling

Before we can model story morals, we need to identify what they are. Fables, defined explicitly as the combination of story and moral (Mandler and Johnson 1977), provide the most natural source. One of the best-known collections of traditional fables are those attributed to Aesop. *The Complete Fables* (Aesop 1998) contains 358 fables, which we grouped according to moral. Fables without a clearly discernible moral were excluded, as were those comprised only of a character reflecting on abstract concepts, because without any events taking place the fable does not really constitute a story. In addition, there are a number of fables in which the moral can only be understood by readers with a prior notion of a specific animal's nature (for example, a snake being 'wicked'); these were also excluded. The result was 85 fables grouped by 27 morals, as listed in Table 1.

Representing Morals

The most intuitive way to represent a story is to model the events that take place. However, the specific events within a story are not directly linked to its moral. For example, the idea of Retribution is that character *A* causes harm to character *B*, which results in *B* causing *A* harm in return. Whether *A* steals from *B*, causes a physical injury or merely offends them doesn't matter, nor does it matter exactly how *B* retaliates. To illustrate why, consider the following storylines:

1. Alice breaks into Bob's house and steals his TV. Bob calls the police and has Alice arrested.
2. Bob catches Alice painting graffiti on his garage door. Bob punches Alice.

Both are examples of Retribution, yet the events taking place are quite different. Thus, to effectively represent morals, we need a foundation other than the events themselves.

Dyer developed Thematic Abstraction Units (TAUs) as a construct for representing the theme or point of a story (Dyer 1983). TAUs are based on characters' planning failures, and have a complex structure which includes information about goals, events and affect (emotion). We aim for a representation using simpler constructs, more in line with Lehnert's *plot units*, which consist of affect states that can be combined to represent particular types of events and plot configurations (Lehnert 1981). However, affect states distinguish

Moral	Fables	Total
Retribution	3, 6, 16, 45, 103, 142 155, 166, 168, 206 242, 244, 270	13
Greed	30, 90, 140, 146, 163, 185, 204, 239, 287	9
Lies	15, 35, 51, 55, 59, 305, 318	7
Realistic Expectations	5, 23, 33, 128, 150, 193	6
Recklessness	38, 40, 68, 181, 301	5
Pride	20, 29, 161, 237	4
Superficiality	37, 76, 102, 119	4
Appreciation	257, 271, 273	3
Hard Work	24, 53, 352	3
Honesty	230, 253, 279	3
Hypocrisy	48, 63, 151	3
Laziness	89, 96, 265	3
Learn from Mistakes	184, 209, 308	3
Unity	71, 86, 215	3
Loyalty	100, 300	2
Planning	241, 336	2
Quality	194, 342	2
Culpability	245	1
Flexibility	143	1
Jealousy	124	1
Overreaction	294	1
Sacrifice	153	1
Selfishness	41	1
Slander	205	1
Temper	58	1
Violence	73	1
Weakness	291	1
TOTAL		85

Table 1: Morals in Aesop's fables

only positive, negative and neutral emotions. We believe a broader set of emotions will yield a representation better suited to the diverse range of morals appearing in stories.

The idea of using character emotions for story generation is not new. The character agents in *Virtual Storyteller* maintain an emotional state based on the OCC theory of emotion (Ortony, Clore, and Collins 1988), which helps determine their behaviour. *Mexica* (Pérez y Pérez 2007) uses emotional links and tensions between characters to develop plot. However, the system uses only two emotions (Love and Hate), and does not attempt to convey morals through the resulting stories. Shaheed and Cunningham (2008) do suggest a relationship between emotions and morals, but their focus is on designing agents rather than modelling story structure.

Modelling Emotion

Modelling emotion is a complex process. Many theories have been proposed over the years (Izard 1977; Plutchik 1980; Russell 1980; Frijda 1986; Lazarus 1991), but few are amenable to computational modelling due to the general informality of the psychological descriptions. One important exception is the OCC theory of emotion (Ortony, Clore, and Collins 1988), which was specifically designed to be computationally modelled. It divides emotions into four categories:

1. **Event-based emotions**, which are experienced as a result of the consequences of events.
2. **Agent-based emotions**, which are directed towards an agent (possibly the self).
3. **Object-based emotions**, which are directed towards objects (either agents or inanimate objects).
4. **Compound emotions**, which are simultaneously agent and event-based. They are felt towards an agent (possibly the self) with respect to the consequences of an event.

The OCC theory is more specific in its definitions than most, and there has been work in formalising the theory using modal logic (Adam, Herzig, and Longin 2009).

We have previously built a system to model character emotion in stories according to the OCC theory (Sarlej and Ryan 2011). Our current implementation covers all 22 emotions the theory describes, though we do not attempt to model the *intensity variables* identified by Ortony, Clore and Collins (1988). We evaluated our model by comparing system-generated emotions for a number of fables to the emotions predicted by people for those same fables. A discrepancy between system output and people's predictions was most apparent for the more complex emotions, but these were generally the same emotions that participants were unable to agree on in their responses; emotions are subjective and ambiguous in nature. The system performed well in the majority of cases.

We model stories as sequences of events and their consequences, a structure which facilitates the eventual goal of story generation. Consequences, in turn, cause characters to experience a variety of emotions. An event can have multiple consequences, each of which can have different emotional effects on the characters involved. Each consequence belongs to an *emotion class*, which captures its emotional effect on an agent. In the OCC theory, the more complex emotions are defined in terms of Joy, Distress, Hope and Fear, as summarised in Table 2. As such, the emotion classes we define correspond to the most basic emotions Joy and Distress (or, in the neutral case, no emotion). Separate predicates, `CausesHope` and `CausesFear`, are used map events to any Hopes or Fears they trigger.

Relief and Disappointment need to be treated differently because, rather than arising as a result of the consequences of events, they occur when a character realises something they hoped for or feared *did not* happen. For example, a character might Hope to find valuables in a safe. If they break into the safe and find it empty, the character should feel Disappointment. To handle this we also map *potential* consequences which did not actually happen (e.g. finding valuables in the safe) to events (e.g. breaking into the safe).

Beliefs About Emotions

Certain emotions depend on other characters' emotions. Handling this requires modelling beliefs. For each emotion, we define a secondary predicate which represents beliefs about other agents' emotions. The emotion definitions in the OCC theory only require one level of beliefs, and hence this is all our implementation provides; an agent can have beliefs about another agent's emotions, but cannot have beliefs about other agents' beliefs. In our current implementation the belief predicates mirror the values of the normal emotion

Emotion	Type	How it is Modelled
Joy	Event-based	a_1 feels Joy if they experience a consequence in the J emotion class.
Distress	Event-based	a_1 feels Distress if they experience a consequence in the D emotion class.
Admiration	Agent-based	a_1 feels Admiration towards a_2 if a_2 carries out an action causing Joy for some other agent.
Reproach	Agent-based	a_1 feels Reproach towards a_2 if a_2 carries out an action causing Distress for some other agent.
Like	Object-based	a_1 will start to Like a_2 if a_1 feels Admiration towards a_2. Equates to the OCC emotion *Love*.
Dislike	Object-based	a_1 will start to Dislike a_2 if a_1 feels Reproach towards a_2. Equates to the OCC emotion *Hate*.
Happy-For	Event-based	a_1 feels Happy-For a_2 if a_1 likes a_2 and believes a_2 experiences Joy.
Pity	Event-based	a_1 feels Pity towards a_2 if a_1 does not dislike a_2 and believes a_2 experiences Distress.
Resentment	Event-based	a_1 feels Resentment towards a_2 if a_1 dislikes a_2 and believes a_2 experiences Joy.
Gloating	Event-based	a_1 feels Gloating towards a_2 if a_1 dislikes a_2 and believes a_2 experiences Distress
Pride	Agent-based	a_1 feels Pride if they believe another agent a_2 (whom they don't dislike) feels Admiration towards them.
Shame	Agent-based	a_1 feels Shame if they believe another agent a_2 (whom they don't dislike) feels Reproach towards them.
Gratification	Event/ agent-based	a_1 feels Gratification if they feel Pride about their own action and experience Joy as a result of one of its consequences.
Remorse	Event/ agent-based	a_1 feels Remorse if they feel Shame about their own action and experience Distress as a result of one of its consequences.
Gratitude	Event/ agent-based	a_1 feels Gratitude towards another agent a_2 if a_1 feels Admiration about a_2's action and experiences Joy as a result of one of its consequences.
Anger	Event/ agent-based	a_1 feels Anger towards another agent a_2 if a_1 feels Reproach about a_2's action and experiences Distress as a result of one of its consequences.
Hope	Event-based	a_1 feels Hope if they experience an event which CausesHope for some consequence.
Fear	Event-based	a_1 feels Fear if they experience an event which CausesFear for some consequence.
Satisfaction	Event-based	a_1 feels Satisfaction if a consequence they hoped for occurs.
Fears-Confirmed	Event-based	a_1 feels Fears-Confirmed if a consequence they feared occurs.
Disappointment	Event-based	a_1 feels Disappointment if a consequence they hoped for does not occur.
Relief	Event-based	a_1 feels Relief if a consequence they feared does not occur.

Table 2: Modelling the OCC emotions

predicates; we assume all agents agree about which emotions a particular agent would feel in response to a particular event.

Location

Our implementation reasons only at the emotional level, and therefore does not consider physical constraints on actions. The only exception is location. For an event involving multiple agents to take place, they must be in the same location. Similarly, for an agent to feel emotions as a result of an event, they must be aware the event took place. We assume all agents in the location where an event occurs are aware of the event; agents in a different location are not. For simple stories where only one event happens at a time it suffices to model two locations (`OnStage` and `OffStage`), which flag whether an agent is present for the current event or not.

Duration of Emotions

In our model we treat most emotions as momentary rather than persistent. That is, an emotion triggered by an event will hold for exactly one time-step following that event. There are two exceptions:

1. **Like and Dislike:** Allowing Like and Dislike to persist lets us model enduring concepts like friendship or enmity.
2. **Hope and Fear:** The consequence an agents hopes for or fears will not necessarily occur (or fail to occur) immediately. For Satisfaction, Disappointment, Fears-Confirmed and Relief to be initiated correctly, Hope and Fear must persist until the relevant consequence either occurs, or explicitly fails to occur.

Generating Emotion Data

To examine the relationship between morals and emotions we require emotion data corresponding to particular morals; this can be produced from Aesop's fables. Morals exhibited by only one or two fables do not provide scope for finding consistent relationships, and thus we used only those morals which appear in at least 4 fables. This leaves us with 48 fables spread across 7 morals: Retribution, Greed, Realistic Expectations, Lies, Recklessness, Pride and Superficiality.

Generating all the emotions occurring in each fable by hand would be both time-consuming and error-prone. Instead, we used our implementation of the OCC theory to automate the task. Each fable was broken down into its constituent events, their consequences, and basic emotional effects (Joy, Distress, Hope and Fear). Running the resulting encoding through our system generated all the emotions felt by each character at every time-point in the story.

Correlating Morals with Emotions

Even stories as simple as fables can involve a large number of emotions. This makes it extremely difficult to draw out common sequences of emotion by hand. We use inductive logic programming, or ILP (Muggleton 1992) for this task. The idea behind ILP is to automatically construct predicate-based rules from supplied background knowledge and examples. We use Aleph (Srinivasan 2003), an ILP system based

Predicate	Description
concurrently_feel (+story,-agent,#emotion,-agent,#emotion)	Emotions felt at the same time by two different agents.
concurrently_feel (+story,-agent,#emotion,#emotion)	Emotions felt at the same time by the same agent.
feels_in_sequence_two (+story,-agent,#emotion,#emotion)	Two emotions felt in sequence by the same agent.
feels_in_sequence_three (+story,-agent,#emotion,#emotion,#emotion)	Three emotions felt in sequence by the same agent.
feels_towards_in_sequence_two (+story,-agent,-agent,#emotion,#emotion)	Two emotions felt in sequence by an agent towards some other agent.
feels_towards_in_sequence_three (+story,-agent,-agent,#emotion,#emotion,#emotion)	Three emotions felt in sequence by an agent towards some other agent.
attitude_towards_in_sequence_two (+story,-agent,-agent,#attitude,#attitude)	A change in attitude of one agent towards another.
concurrently_feel_towards (+story,-agent,#emotion,-agent,#emotion)	Emotions felt by two agents towards each other at the same time.
concurrently_feel_towards_in_sequence_two (+story,-agent,#emotion,#emotion,-agent,#emotion,#emotion)	Emotions felt in sequence by two agents towards each other.

Table 3: Predicates provided as background knowledge for Aleph to use in constructing rules

on Progol. We apply Aleph's tree-learning approach to build a classification tree which, given the emotion data for a fable, attempts to predict which moral the fable conveys. As input for tree-learning, Aleph requires background information, which consists of the emotions taking place in each fable along with definitions of predicates to use for constructing rules, and examples mapping fables to morals.

Predicate Definitions

The predicates for Aleph to use for rule construction had to be defined by hand. Our goal was to establish a relationship between morals and emotions, without relying on specific events or their consequences. Thus the predicates we defined focus on the emotions taking place within the story. A listing of the predicates used, along with a brief description of each, is provided in Table 3. As per standard Progol syntax, the name of an argument represents its type. The prefixes specify whether Aleph should substitute those arguments with variables (+ or -, for input and output variables respectively) or constants (#) during its search for rules. Note that we make a distinction here between `emotion` and `attitude`. In this case, `emotion` includes all the OCC emotions except Like and Dislike. Since Like and Dislike persist over time, finding that an agent feels Like and Like in sequence is not very meaningful. On the other hand finding that an agent feels, for instance, Joy and Joy in sequence is meaningful, because the agent must have experienced two separate events resulting in Joy.

Approach to Validation

Due to the small number of examples available it was not feasible to separate the data into disjoint training and test sets. Instead, we use leave-one-out cross-validation to estimate the accuracy of our model, maximising the number of examples that can be used for training each fold. We partitioned the data into a training set of 47 examples and a test set consisting of the one remaining example. This was repeated 48 times, with a different example left out of the training set each time, allowing us to estimate the overall accuracy of the classification on unseen examples.

Experiment 1: Learning Individual Morals

Our original goal was to learn rules for all 7 morals, but the resulting classifier performed poorly. Table 4 shows the confusion matrix produced from the cross-validation results. The only moral with reasonable recall and precision is Retribution. Given the small number of fables available for some morals, this outcome is not surprising. Retribution, the most successfully predicted moral, also had the most examples available. It is likely there was simply not enough data to isolate characteristic emotion sequences for the other morals. However, patterns emerged from the confusion matrix which could be exploited; some morals were consistently misclassified as certain other morals.

Experiment 2: Learning Groups of Morals

The fact that certain morals were consistently confused with one another could indicate an inherent similarity. To investigate this possibility we grouped together those morals that were frequently confused, as follows:

- Retribution and Lies (20 fables)
- Greed and Realistic Expectations (15 fables)
- Recklessness, Pride and Superficiality (13 fables)

This partitions the data into only three classes, but with significantly more examples of each. We used the same cross-validation approach, and the results were more promising. Table 5 presents the relevant confusion matrix.

Intuitively, the similarity between Retribution and Lies makes sense. To demonstrate that lying is bad, fables present

	Retribution	Lies	Greed	Realistic	Recklessness	Pride	Superficiality	RECALL
Retribution	9	1	1	0	0	2	0	69.2%
Lies	3	2	1	0	1	0	0	28.6%
Greed	0	1	2	4	1	1	0	22.2%
Realistic	0	0	6	0	0	0	0	0.0%
Recklessness	0	1	3	0	0	0	1	0.0%
Pride	1	0	1	0	0	1	1	25.0%
Superficiality	1	1	1	1	0	0	0	0.0%
PRECISION	64.3%	33.3%	13.3%	0.0%	0.0%	25.0%	0.0%	

Table 4: Confusion matrix for individual morals (Experiment 1)

	Retribution/Lies	Greed/Realistic	Recklessness/Pride/Superficiality	RECALL
Retribution/Lies	17	2	1	85.0%
Greed/Realistic	1	13	1	86.7%
Recklessness/Pride/Superficiality	3	3	7	53.8%
PRECISION	81.0%	72.2%	77.8%	

Table 5: Confusion matrix for grouped morals (Experiment 2)

the negative consequences of a lie for another character, followed by a negative consequence for the liar as a punishment. This is just like the pattern described earlier for Retribution. Greed is depicted by characters who want more than they have. In attempting to get it they not only fail, but often lose what they originally had. Fables expressing the importance of Realistic Expectations focus on characters attempting something outside their abilities, and similarly failing. It is easy to imagine the resulting emotions would be comparable. The recall for Recklessness/Pride/Superficiality is low, but the precision is fairly high. Evidently these three morals are not very similar to each other, but are distinctively different from the other four.

Figure 1 shows the classification tree constructed by Aleph using all 48 fables for training. The numbers below the leaf nodes show how many fables of that class were correctly predicted by that branch of the tree. Three examples of Superficiality/Pride/Recklessness and one of Lies/Retribution are not correctly predicted by this tree, and thus the numbers below those classes do not sum to the total number of fables in each. It should be noted that the specific rules produced by Aleph depend on both the model of emotion chosen and the predicate definitions provided, which limit the relationships Aleph can learn. A wider range of predicates could potentially yield a better model.

Future Work

There are several possible avenues for future work stemming from this research. Although our results suggest a relationship between morals and emotions, they do not describe this relationship in enough detail, or with sufficient accuracy, for practical applications. More reliable results could be achieved by increasing the number of examples for each moral, obtained from other collections of fables and folk tales (Harris 1955; Cole 1983; Yolen 1986; Forest 1996). A broader dataset would not only help identify relationships between emotions and individual morals, but also allow a wider range of morals to be investigated.

The nature of Aleph's tree-learning is to construct the simplest possible classification tree for the supplied examples. This approach produces rules that may not incorporate all the emotion sequences relevant to a particular moral. Information that could be important for a human reader's interpretation of that moral may be left out or pruned from the tree. It may be worth investigating other learning approaches, both within Aleph (for example, standard ILP rule-learning, one class at a time) and using other systems.

Regardless of how accurately a model based on emotion predicts the moral of an unseen fable, it is impossible to conclude whether those emotions are *sufficient* to convey that moral, or only a contributing factor. There may be other features of the fable which are necessary for a reader to understand its message. The only way to determine whether appropriate patterns of emotion are adequate to convey morals is to build a system which works in the opposite direction; it should generate stories that conform to particular patterns of emotion, which can then be evaluated by people. This is the direction we intend to pursue.

Conclusion

In this paper we proposed a relationship between the moral of a story and the patterns of emotion experienced by the characters within that story. Due to the small number of examples available, we were unable to reliably identify emotions corresponding to individual morals. Nevertheless, a relationship between emotions and morals is evident from our results when similar morals are grouped together. We believe a larger pool of examples would lead to significantly better results and more specific correlations. It remains to be seen whether these relationships will be sufficient to generate stories with given morals; that is an area for future work.

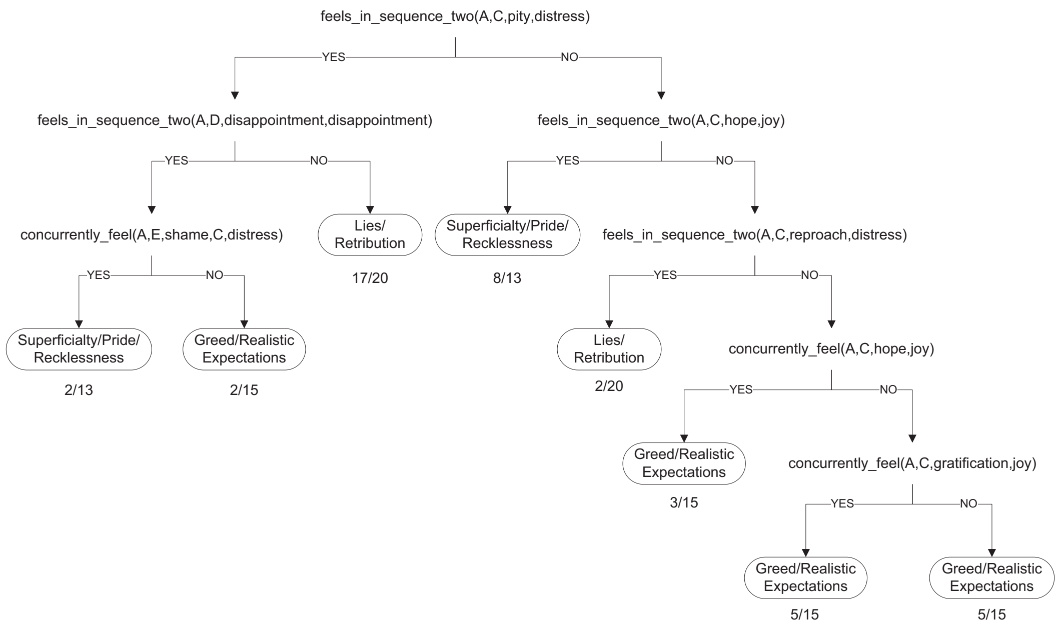

Figure 1: Moral classification tree, generated from the complete dataset of 48 fables

References

Adam, C.; Herzig, A.; and Longin, D. 2009. A logical formalization of the OCC theory of emotions. *Synthese* 168(2):201–248.

Aesop. 1998. *Aesop: The Complete Fables*. Penguin Classics. London, UK: Penguin Books. Translated by Olivia and Robert Temple; introduction by Robert Temple.

Cheong, Y.-G., and Young, R. 2008. Narrative Generation for Suspense: Modeling and Evaluation. *Interactive Storytelling* 144–155.

Cole, J. 1983. *Best-Loved Folktales of the World*. New York: Anchor Books.

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

Forest, H. 1996. *Wisdom Tales from Around the World*. Atlanta, Georgia, USA: August House.

Frijda, N. H. 1986. *The Emotions*. New York: Cambridge University Press.

Harris, J. C. 1955. *The Complete Tales of Uncle Remus*. Boston: Houghton Mifflin.

Izard, C. E. 1977. *Human Emotions*. New York: Plenum Press.

Lazarus, R. S. 1991. *Emotion and Adaptation*. New York: Oxford University Press.

Lehnert, W. G. 1981. Plot Units and Narrative Summarization. *Cognitive Science* 5(4).

Mandler, J., and Johnson, N. 1977. Remembrance of Things Parsed: Story Structure and Recall. *Cognitive Psychology* 9(1):111–151.

Muggleton, S. 1992. *Inductive Logic Programming*, volume 38 of *The APIC Series*. London: Academic Press Limited.

Ortony, A.; Clore, G. L.; and Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

Pérez y Pérez, R. 2007. Employing emotions to drive plot generation in a computer-based storyteller. *Cognitive Systems Research* 8(2):89–109.

Plutchik, R. 1980. *Emotion: A Phsychoevolutionary Synthesis*. New York: Harper & Row.

Russell, J. A. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39(6):1161–1178.

Ryan, K. 1991. The Narrative and the Moral. *The Clearing House* 64(5):316–319.

Sarlej, M., and Ryan, M. 2011. A Discrete Event Calculus Implementation of the OCC Theory of Emotion. In *The 4th Workshop on Intelligent Narrative Technologies (INT4)*. Stanford University, Palo Alto, California: AAAI Press.

Shaheed, J., and Cunningham, J. 2008. Agents making moral decisions. In *Proceedings of the ECAI08 Workshop on Artificial Intelligence in Games*.

Srinivasan, A. 2003. The Aleph Manual. Technical report, Computing Laboratory, Oxford University.

Theune, M.; Faas, S.; Nijholt, A.; and Heylen, D. 2003. Emotional Characters for Automatic Plot Creation. In *Technologies for Interactive Digital Storytelling and Entertainment*.

Yolen, J. 1986. *Favourite Folktales from Around the World*. New York: Pantheon Books.