In Proceedings of the Sixth International Conference on Cognitive Modeling, 392-393. Mahwah, NJ: Lawrence Earlbaum.

A Computational Model for Acquisition of Counterintuitive Concepts

M. Afzal Upal
University of Toledo,
Toledo, OH, 43606
afzal.upal@eng.utoledo.edu

Introduction and Background

In a seminal study, Bartlett (1932) asked British university students to read a culturally unfamiliar Native North American folk tale called "the war of the ghosts" and retell it to others in writing who in turn passed it on to others. Over several generations of retellings culturally unfamiliar concepts were distorted and replaced by more familiar concepts; for instance, a canoe was replaced by a rowboat. Bartlett reasoned that unfamiliar concepts such as canoe are more difficult to represent and thus less likely to be represented and transmitted than familiar concepts. However, recent studies following up on Bartlett's work (Boyer and Ramble 2001, Barrett and Nyhof 2001, Atran 2003) have drawn the seemingly opposite conclusion, namely that under some conditions, counterintuitive concepts (i.e., concepts that violate intuitive expectations) are actually better recalled than relatively more intuitive concepts. Barrett and Nyhof (2001) asked their subjects to remember and retell three of six different Native American folktales containing an equal number of intuitive and counterintuitive concepts. A content analysis of what they remembered showed that subjects remembered significantly larger number of counterintuitive concepts than intuitive concepts. In subsequent experiments, Barrett and Nyhof (2001) and Boyer and Ramble (2001) constructed a number of stories containing an equal number of three different types of concepts: the minimally counterintuitive concepts (i.e., concepts that violate intuitive expectations regarding one or two feature values) such as "furniture that flies in the air", ordinary concepts such as "furniture that stays where you put it", and maximally counterintuitive concepts (i.e., concepts that violate intuitive expectations regarding multiple feature values) such as "furniture that flies in the air, melts when try to grab it, and is made of uranium." When subjects from a variety of cultural backgrounds were asked to recall the stories containing an equal number of three types of concepts, they recalled a significantly larger number of minimally counterintuitive concepts than both ordinary and maximally counterintuitive concepts. The results regarding better recall for minimally counterintuitive concepts visa-vi ordinary concepts are in accordance with a host of studies showing that unusual or distinctive stimuli are generally better remembered than stimuli that are not (see Waddill and McDaniel 1998 for a review). A number of factors have been proposed to account for these effects including:

- more attention being paid to the unusual stimuli resulting in richer spontaneous elaboration and encoding for them and less attention and time being spent on processing the usual stimuli,
- unusual stimuli being encoded in their own category, different from the one into which the common stimuli are clustered,
- unusual stimuli being distinctive in the retrieval set that is formed.
- Boyer and Barrett have argued that it is the economy of representation as simple negation of a feature value and the ability to invoke the counterintuitive concept to make predictions because of its set of non-violated feature values that makes minimally counterintuitive concepts easier to recall than ordinary concepts. They argue that maximally counterintuitive concepts are harder to recall because they have little "inferential potential" as they violate too many intuitive expectations.

There are two major problems with the previous explanations. First, the proposed mechanisms are little more than descriptive accounts that do not really explain as to why people use these concept acquisition and memory mechanisms as opposed to using different concept acquisition and memory mechanisms. Second, they do not explain the reasons for the contradictions among findings of various studies including Bartlett's original study and more recent work that seems to confirm findings regarding better recall for ordinary concepts (Atran 2003). Instead of embedding concepts in the structure of a story, Norenzayan and Atran (2003) simply asked their subjects to remember and recall an equal number of intuitive, minimally counterintuitive, and maximally counterintuitive concepts. Unlike Barrett *et al.* and Boyer *et al.*'s studies, Norenzayan *et al.* observed that intuitive concepts were better recalled than minimally counterintuitive concepts which were better recalled than maximally counterintuitive concepts.

This paper proposes an opportunistic (Francis 1995) concept learning framework that can account for the seemingly contradictory findings. We are currently working to implement the proposed technique in a machine learning system called CICL (CounterIntuitive Concept Learner).

Knowledge-driven Concept Learning

The idea that concept learning is strongly influenced by theoretical and causal knowledge that people possess is not new (Murphy & Medin 1985, Kunda, Miller, & Claire 1990, Rehder 1999). When exposed to a new object belonging to a novel concept, people use their prior knowledge about similar concepts to generate expectations about the new concept. If the new concept is grossly incompatible with their expectations, it is relatively more difficult to acquire than a concept that is in accordance with expectations. Murphy and Wisniewski (1994) reported that when features of a concept are consistent with prior knowledge people found categories easier to learn compared to situations in which features do not fit in with prior knowledge. Pazzani (1991) and Wattenmaker (1995) found that people learned a concept in fewer trials if its definition was congruent with prior knowledge. Heit (1998) found that when subjects were allowed to self pace during a training phase in which they were presented with intuitive concepts (such as "shy person who rarely attends parties") and expectation violating concepts (such as "shy person who attends parties often"), they spent significantly more time processing expectation violating concepts than the time spent processing expectation compliant concepts.

There is also evidence that part of the extra time taken to process expectation violating concepts is spent on trying to find answers to the questions raised by the expectation violations. When Kunda *et al.* (1990) presented subjects with expectation violating concepts such as a blind marathon runner, Harvard educated carpenter, and feminist bank teller they attempted to answer questions such as how could a Harvard educated person become a carpenter? using their prior knowledge about Harvard educated people, about carpenters, and about people in general. Kunda *et al* (1990) report that subjects "had no trouble combining sometimes incongruous social categories. There were no cases in which subjects reported difficulty or failed to complete the task… some subjects spontaneously provided causal narratives that seemed designed to address the question of how a member of one category acquired membership in the other."

An Opportunistic Model of Concept Forgetting

An agent with finite memory resources cannot possibly remember every object it perceives (Altman & Gray 2002). An opportunistic agent should prefer to forget those concepts that require less cognitive processing to acquire. For a rational agent whose goal is to increase its knowledge of the environment to improve its problem solving performance, it makes sense to forget objects that are expectation compliant at a higher rate because they require minimal effort to predict. Similarly, for an agent living in a world full of uncertainty, it makes sense to preferably forget bizarre objects whose occurrence it cannot explain.

We hypothesize that human memory forgets concepts at different rates depending on how much cognitive effort was required to acquire those concepts. We believe that this differential rate of forgetting accounts for the differences between recall rates of ordinary concepts, recall rates of minimally counterintuitive concepts, and recall rates of maximally counterintuitive. The concepts that do not violate any expectations require minimal cognitive effort and hence are most easily forgotten. When presented with minimally counterintuitive concepts people attempt to answer questions such as "why would someone design a piece of furniture that can fly?" What question people ask depends on the concept. For instance, in case of artifact concepts the crucial question is the functionality of the artifact leading to the question why would someone design an artifact with these properties. People learn new concepts as a result of being able to answer such questions. These newly learned concepts are ranked lower for forgetting because of the cognitive effort spent in learning them. The maximally counterintuitive concepts require more cognitive processing than ordinary concepts because they violate expectations triggering the explanation seeking mechanism but the search is quickly abandoned when no explanation can be formed.

Our hypothesis, if true, would also explain the results obtained by Norenzayan and Atran (2003) that show that ordinary concepts are recalled better than counterintuitive concepts when presented without the structure of a story. The prior knowledge that a person possesses impacts the creative explanation formation process. For instance, when Heit (2001) provided subjects with verbal arguments such as "shy people may also go to parties in an attempt to overcome their shyness and to meet new people" the impact of expectation violation disappeared. We hypothesize that stories in conjunction with people's common knowledge about stories act as verbal arguments by providing a context that explains the occurrence of counterintuitive concepts such as "a sobbing oak." When stripped of this context, as in the experiments performed by Norenzayan and Atran (2003) the minimally counterintuitive concepts become maximally counterintuitive concepts and can no longer be justified. However, such concepts can be justified in the context of stories which leads to better recall for them in that context.

Conclusion

Previous models of memory and concept learning are unable to explain differences between recall rates of moderately expectation violating concepts, the recall rates of ordinary concepts, and the recall rates of maximally expectation violating concepts in various situations. This paper has presented a model based on Bayesian belief networks that explains the seemingly contradictory results of the free recall experiments conducted with human subjects. Our theory also gives rise to a number of testable predictions. Plans are in the works to conduct experiments with human subjects to confirm implications of our theory. We are also readying CICL, a computer system designed using the concept learning and memory system, for public release.

Acknowledgement: Thanks to Ryan Tweney, Jason Slone, Nicholas Jutte, and David Mullen for guidance and suggestions.

References

Atran, S. (2003), *In Gods We Trust*, Oxford Univ Press.

Barrett, J. & Nyhof, M. (2001). Spreading non-natural concepts. *Journal of Cognition and Culture*, 1, 69-100.

Boyer, P. & Ramble, C.. (2001). Cognitive templates for religious concepts. *Cognitive Science*, 25, 535-564.

Murphy, G. (2002) The Big Book of Concepts. MIT Press.

Waddill, P. J. & McDaniel, M. A. (1998) Distinctiveness effects in recall, *Memory and Cognition*, 26, 108-120.