Truth versus Useful Lies

Daniel S. Levine
University of Texas at Arlington

Abstract: A religious leader in a Kurt Vonnegut novel advocated abandoning the search for truth in favor of a search for useful lies. The switch from truth to useful lies is a good metaphor for what typically occurs in the development of human decision making. As we grow from childhood to adulthood, we gradually make decisions less on the basis of the literal truth of events (what the psychologists Reyna and Brainerd call *verbatim encoding*) and more on the basis of categories of experience (what those psychologists call *gist encoding*). The switch from verbatim to gist makes decisions fit into patterns which are efficient and enable us to live in a complex society. Yet it also leads to some characteristic distortions, such as incorrect encodings of probabilities.

I will sketch some brain processes that may be involved in the "fuzzy traces" that lead to gist encoding and distortion of probability weights. One part of the frontal lobes, the orbitofrontal cortex, plays the major role in making us live by our gists, for good and for ill. But we also need something like the child in *The Emperor's New Clothes* to stand apart from our culture and tell us when we need to restore our verbatim encoding, that is, to seek truth instead of useful lies. Two other parts of the frontal lobes seem to play the roles of challenger (the anterior cingulate cortex) and truth-teller (the dorsolateral prefrontal cortex).

Truth and Lies: Verbatim and Gist Representations

In Kurt Vonnegut's novel *Cat's Cradle* (Vonnegut, 1962), a political leader and a religious leader take over the unproductive, poverty-stricken Caribbean country of San Lorenzo and set out to change its social mores. The religious leader, Bokonon, decides that acceptance of poor conditions requires an elaborate fantasy life, much of it revolving around the myth of persecution of the "saintly" Bokonon by the "tyrannical" political leadership. After all, he concludes, every religion including his own is largely based on lies. In his own "Books of Bokonon" he writes: "Live by the harmless untruths that make you brave and kind and healthy and happy."

Vonnegut's advocacy of "useful lies" is grounded in a cynical outlook on human social organization: his lies are intended as a mechanism for coping with unpleasant truths about one's environment. But could they also be a useful metaphor for much of human and machine information processing? Starting with humans, how much of our day-to-day functioning is based on distortions of the truth about the world? And is Vonnegut's character Bokonon correct in labeling such distortions "harmless"?

We will start with some psychological studies of memory and decision making. Two cognitive psychologists, Charles Brainerd and Valerie Reyna (e.g., Brainerd & Reyna, 1990; Reyna & Brainerd, 1995; Reyna, Lloyd, & Brainerd, 2003) have noted experimental findings in both memory and decision making that suggest our processing, storage, retrieval, and/or reasoning from external information deviates from the veridical in systematic ways. In the case of memory, for example, experimental participants exposed to a list of words all related to, but not including, a key word (e.g., a list that includes "dream," "bed," "pillow," "sheets" and "nightmare" but not "sleep") will be more than likely to falsely remember having seen the key word. In decision making under risk, participants tend to make different choices between identical alternatives when those alternatives are framed in terms of gains or losses. For example, if faced with an epidemic expected to kill 600 people, they prefer the certain saving of 200 to a 1/3 probability of saving 600 with a 2/3 probability of saving none, but prefer a 2/3 probability of all 600 dying with a 1/3 probability of nobody dying to the certainty of 400 dying.

While these phenomena are often regarded as signs of human irrationality, Brainerd and Reyna see them as simply the "dark side" of a sophisticated cognitive system that enables us to process and act on the essentials of information while ignoring irrelevant details. Their *fuzzy trace theory* posits that we possess two somewhat separate systems for encoding information: literal or *verbatim* encoding, and intuitive or *gist* encoding. Verbatim encoding means literal storage of facts, whereas gist encoding means storing the essential intuitive meaning or "gist" of a situation. As people develop from childhood to adolescence to adulthood, there is a gradual switch from greater reliance on verbatim encoding to greater reliance on gist encoding. This can lead to inaccuracies but also enhance the ability to categorize a situation as similar to others that have occurred before. Such categorizations are the source of heuristics that can both simplify decision making and make it less precise. As Reyna et al. (2003) state:

> Although development involves increasing resistance to interference and greater reliance on gist, mature processing has particular pitfalls. However, reliance on gist allows mature reasoners to perceive the underlying patterns across superficially disparate problems, which is essential to achieving cognitive consistency – a fundamental criterion of rationality (Tversky & Kahneman, 1986). Thus, gist-based reasoning provides global advantages to performance, despite local pitfalls in solving particular problems. (Reyna et al., 2003, p. 232).

In other words, when we are children, or when we enter a new domain (such as learning a new language or a new skill), we tend to store and remember things we see and hear literally, without necessarily understanding their meaning. As adults or experts, we get an intuitive "feel" that enables us to process the environment in a more meaningful and selective manner. Our adult/expert gist memories are more robust and less vulnerable to disruption than our childhood/novice verbatim memories.

Yet the switch from verbatim to gist has its costs. And we often make errors because we encode the wrong gist or use a gist incorrectly. One example is the misestimation of risk that often occurs in decisions by physicians. For example, if 90 percent of women with breast cancer have positive mammograms, physicians seeing a positive mammogram may overestimate the likelihood of cancer, which also (by Bayes' rule) depends on the base rate of cancer in the population being studied (for a review see Krynski & Tenenbaum, 2007). Another example is acting on social stereotypes. We often learn that someone is, say, an African-American, a Chinese, a mathematician, or a lawyer, and interpret their behavior in terms of the "gist" of what we have learned about that category of people. In the process we may overlook aspects of that person's actions that do not fit some of the attributes we ascribe to the category.

The most creative ways of being and thinking, then, require active use of *both* the gist and verbatim systems. As Brainerd and Reyna point out, we need the gist system to understand overall trends, and to see analogies between different problems or domains that superficially appear very different. Yet we also need the verbatim system to override inaccurate determinations of the gist system in new, unfamiliar, or atypical contexts. This is why the most creative people are often described as having a "childlike" quality. They have a kinship with the little boy in the fairy tale *The Emperor's New Clothes*, who hears all the adults praising the (nonexistent) clothes their ruler is wearing and points out "but he has nothing on" (Andersen, 2007). Creative people also have a kinship with ideals of reclaimed innocence from many religions. For example, the Mahayana branch of Buddhism aspires to a state called Nirvana wherein we perceive things just as they are, without the aid of categories or labels. In this state:

> ... there is nothing but what is seen of the Mind itself; where there is no attachment to external objects, existent or non-existent; where ... there is an insight into the abode of reality as it is; where, recognising the nature of mind in itself, one does not cherish the dualism of discrimination ... (Burtt, 1955, p. 162)

How might the gist and verbatim systems be represented in the higher cognitive levels of our brains? What is the mechanism by which we learn rules in the course of development, under the heavily influence of social expectations? And what is the mechanism by which we selectively override our developmental conditioning when the situation calls for it?

The Brain Executive System and Rules

The author has been engaged in a long-term research program to develop neural network theories of the involvement of different parts of the brain's executive system in simultaneously encoding multiple types of rules: not just verbatim versus gist but also competitive versus cooperative (Eisler & Levine, 2002); creative versus rigid (Levine, 2005); and deliberative versus heuristic (Levine, 2007; Levine & Perlovsky, 2008). More details about brain regions are contained in those articles but we can summarize some of them here.

The movement from predominantly verbatim to predominantly gist encoding is related to maturation of the brain's frontal lobes, which occurs gradually from birth to young adulthood (Thompson et al., 2000). Within the frontal lobes there are three executive areas that have been found to have specific and dissociable functions: the *orbitofrontal cortex* (OFC), *anterior cingulate cortex* (ACC), and *dorsolateral prefrontal cortex* (DLPFC).

The OFC creates links between the sensory and association areas of cortex and subcortical emotional centers such as the amygdala and hypothalamus, and thereby stores positive and negative emotional valences for different stimuli, concepts, and potential behaviors (Öngür & Price, 2000; Rolls, 2000). Human patients with damage to the OFC are severely impaired in their decision making process due to the detachment of their cognitive functions from their emotional connections, which can lead to either impulsiveness or excessive deliberation (Damasio, 1994). This area seems a likely candidate for the encoding of many of the gists of common contexts and situations that we learn throughout development, particularly from our families and cultures.

Yet the OFC and its connections to both sensory and emotional areas do not tell us unambiguously what rules of action to follow. Sometimes we are aware of a conflict between an “easier” and a “harder” rule or course of action. Another frontal lobe region, the ACC, is particularly sensitive to the presence of a possible rule conflict. For example, DeNeys, Vartanian, and Goel (2008) told their participants that a certain group of people consisted of some number of lawyers and some number of engineers: the numbers could either be balanced (500 lawyers and 500 engineers) or imbalanced (995 lawyers and 5 engineers, or the reverse). A hypothetical person was identified as a member of that group and given a description that fits a common stereotype of one of the two professions; then the participant was asked to determine the probability of that person being a lawyer or an engineer. DeNeys et al. applied functional MRI to their participants’ brains and found greater activation of the ACC (conflict detector) in the incongruent case (when, for example, a stereotypical description of an engineer was combined with a statement that 995 out of the 1000 were lawyers) than in the congruent case. In another fMRI study that involved rewards and punishments, Blair et al. (2006) found that the ACC was more active when a subject had to choose between two undesirable alternatives than between two desirable alternatives, the reverse of the activation pattern for the OFC.

In the imaging experiment of DeNeys et al. (2008), awareness of incongruity did not always lead to making a correct probability judgment, one that takes into account not only the stereotypical description but also the percentage of lawyers and engineers in the total population. The commonest response is to ignore the population profile, which is called *base rate neglect* (Kahneman & Tversky, 1973). DeNeys et al. found that those subjects who did not show base rate neglect had greater activation of the third major executive subdivision of the frontal lobes, the DLPFC. The DLPFC has been known for a long time to be involved in complex working memory manipulation, so is the brain region most involved in complex logical deliberation (Fuster, 2000).

Hence, using the gist/verbatim distinction of fuzzy trace theory, we can summarize the functions of these three highly interconnected executive brain regions as follows:

OFC: learning and enhancing gists
ACC: challenging prevailing gists
DLPFC: determining verbatim overrides, or new gist overrides

Of course, because of the interconnections between these three areas and many others (the sensory, motor, and association areas of cortex; the corpus striatum, which implements actions; and the emotional regions of the amygdala, hypothalamus, and midbrain), any neat parcellation of area functions is simplistic. Yet it is useful to see how these sub-functions might interact in a network. The interactions among these brain areas also provide guidance for any intelligent system that needs to act with different degrees of creativity or automaticity in different contexts. In the words of Reyna et al. (2003), human reasoners (or, by extension, any machines that have a comparable level of intelligence)

> … shift their level of representation, called task calibration, to adjust for such factors as response mode (more precise responses generally require more precise representations), tolerance for error in a the particular situation, and the ambiguity of inputs (e.g., the usual variation of numerical scores along a dimension.) (p. 214).

Task calibration is one of the great unsolved meta-problems of (brain-based) neural network theory. Yet suitable combinations of network architectures developed for other problems – one example being the *adaptive resonance theory* (ART) networks (Carpenter & Grossberg, 1987; Carpenter, Grossberg, & Reynolds, 1991), which can categorize inputs at any desired level of precision or vagueness and can learn the environments' regularities and irregularities -- seem likely to solve it in the near future. We earnestly hope that human beings who learn and apply the principles of task calibration can end up braver, kinder, healthier, and happier than the poor ignorant people of Vonnegut's San Lorenzo!

References

Andersen, H. C. (2007). *Fairy Tales* (trans. M. H. Hult). New York: Barnes and Noble Classics.

Blair, K., Marsh, A. A., Morton, J., Vythilingam, M., Jones, M., Mondillo, K., Pine, D. C., Drevets, W. C., & Blair, J. R. (2006). Choosing the lesser of two evils, the better of two goods: Specifying the roles of ventromedial prefrontal cortex and dorsal anterior cingulate in object choice. *Journal of Neuroscience*, *26*, 11379-11386.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*, 3-47.

Burtt, E. A. (Ed.) (1955). *The Teachings of the Compassionate Buddha*. New York: Mentor.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, *37*, 54-115.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, *4*, 565-588.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Grosset/Putnam.

DeNeys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brain detects we're biased. *Psychological Science*, *19*, 483-489.

Eisler, R., & Levine, D. S. (2002). Nurture, nature, and caring: We are not prisoners of our genes. *Brain and Mind*, *3*, 9-52.

Fuster, J. M. (2000). The prefrontal cortex of the primate: A synopsis. *Psychobiology*, *28*, 125-131.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237-251.

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, *136*, 430-450

Levine, D. S. (2005). Angels, devils, and censors in the brain. *ComPlexus, 2*, 35-59.

Levine, D. S. (2007). How does the brain create, change, and selectively override its rules of conduct? In R. Kozma & L. Perlovsky (Eds.), *Neurodynamics of Higher-level Cognition and Consciousness* (pp. 163-181). Heidelberg: Springer-Verlag.

Levine, D. S., & Perlovsky, L. I. (2008). Simplifying heuristics versus careful thinking: scientific analysis of millennial spiritual issues. *Zygon*, *43*, 797-821.

Öngür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys, and humans. *Cerebral Cortex*, *10*, 206-219.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1-75.

Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative theory of judgment and decision making. In S. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Making* (pp. 201-245). New York: Cambridge University Press.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, *10*, 284-294.

Thompson, P. M., Giedd, J. N., Woods, R. P., MacDonald, D., Evans, A. C., & Toga, A. W. (2000). Growth patterns in the developing brain detected by using continuum mechanical tensor maps. *Nature*, *404*, 190-193.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *39*, 252-278.

Vonnegut, K. (1962). *Cat's Cradle*. New York: Henry Holt and Company.