Computational Humor: Promises and Pitfalls

John Charles Simon

John Charles Simon Consulting
jsimon.wwl@gmail.com

Abstract

Creating an AI device that is both easy to control and comfortable to interact with will likely require algorithms for accurately interpreting conversational speech. Homonyms and homophones represent a particular challenge in this regard, thus the study of puns and other forms of humorous wordplay can be informative. Moving beyond the simple resolution of word uncertainty to an understanding of humor is, however, problematic. The Mutual Vulnerability Theory of Laughter identifies numerous variables involved in our differentiating humorous and nonhumorous stimuli. These include available information, type and degree of relationship with others, personal history, culture, and even mood. It also suggests there will be potential liabilities for AI users, retailers, and developers resulting from even successful attempts to identify, respond to, and create humor, as all require the highlighting of vulnerabilities.

Humor as Guide and Goal

As a greater percentage of the population encounters a growing number of technologies exploiting artificial intelligence, we should anticipate overall productivity to increase substantially. However, as with any technology, advanced capabilities typically generate higher expectations on the part of operators. For many AI devices, a user's preferred command interface will be his or her native spoken language, this despite the difficulties existing speech-recognition algorithms exhibit when attempting to correctly interpret identical- or similar-sounding words (or word combinations) with different meanings.

For this reason, the growing interest of the AI community in computational humor is well-justified. Many forms of verbal humor, puns for example, rely on atypical or unexpected juxtapositions of homonyms and homophones. The development of programs that correctly decipher various types of humorous wordplay will inevitably facilitate more accurate interpretations of conversational speech.

Beyond this basic word recognition problem is the potential for humor appreciation (and perhaps creation) by AI

devices. These capabilities, too, would undoubtedly make interactions between users and devices more familiar, more informal, and thus more enjoyable.

Innovations in this realm will not, however, come without risk. Comedians, cartoonists, and politicians, among others, know that humor is highly subjective. Witticisms might be understood and welcomed by some but not others. Suggestions that are obviously ridiculous to one group may be seen as callous or highly offensive by another. And jokes that are appropriate at one point in time or in one social circumstance are often unsuitable in others. Failure on the part of AI programmers to understand these, often subtle, qualities of humor will undoubtedly result in negative outcomes for users, retailers, and developers.

Unfortunately, established theories typically employed to understand humor are generally acknowledged to be, at best, incomplete (Morreall 1986). Superiority Theory, Incongruity Theory, Tension Relief Theory, each conceived in ancient Greece (Perks 2012), as well as their more modern derivatives, all exhibit gaps, inconsistencies, and contradictions.

In this paper I will briefly summarize a relatively new and comprehensive theory to explain humor (and laughter) and, using the insights it provides, identify potential liabilities of which the AI community should be aware.

Mutual Vulnerability Theory

Hypotheses to explain humor consistently run into the same problem. Humor is so varied in its form, content, and effect that no single "formula" can describe it. Cognitive models in particular, typically variants of the Incongruity Theory favored by most researching verbal humor, tend to ignore the seemingly banal constituents of most conversational humor (Provine 1993). Superiority Theory and Tension Relief correctly seek to explain laughter first, but are both self-contradictory.

In the mid-1990s, I developed a new conceptual model to explain laughter and, by extension, humor (Simon 2008). Like Superiority Theory, the Mutual Vulnerability Theory (MVT) holds that laughter is fundamentally a form of communication, not simply a response to uncertainty,

inconsistencies, or their favorable resolution. However, rather than an expression of one's superiority, MVT defines laughter as *a vocal affirmation of mutual vulnerability*. Roughly translated, laughter says, "I'd like to remind you that we share some degree of limitation, frailty, or failing." For the purposes of explaining laughter, we should think of vulnerability as a trait (a physical characteristic, behavior, or combination thereof), or the absence of a trait, that makes achieving a goal less probable than "normal," but not so improbable as to constitute a "deficiency." The desire to express this sense of shared vulnerability is typically what we refer to as "amusement."

Laughter is adaptive. We can deduce an evolutionary history going back at least 16 million years, since all the Great Apes have laugh-like vocalizations and use them in the some of the same circumstances humans do (Fry 1977, Fossey 1983, van Lawick-Goodall 1971). These include tickling, play chases, rough-and-tumble play fighting, all of which highlight physical vulnerabilities. So important is laughter that it can be expressed by an individual tens- or even hundreds-of-thousands of times in a lifetime. It serves to minimize the degree or importance of perceived status differences in a way that promotes play (critical for learning, creativity, and exploration), cultivates and strengthens interpersonal relationships, and regulates behavior.

The greatest strength of MVT is its ability to explain variation in the human laugh response. Individuals with different priorities, ages, personalities, knowledge bases, experiences, etc. will differ in what they perceive to be a vulnerability. Furthermore, as with most forms of communication, the desire to express any emotion will additionally depend on the prior relationship between sender and receiver(s), their moods and modes, the larger social environment (e.g., family, region, sociopolitical group, religion), situational context (e.g., birthday party vs. funeral), and so on.

Simple forms of nonverbal communication, even though they convey a single message, can have different meanings depending on the context and the relationship between participants. For example, the message of your car horn is basically "hey," and yet a toot on your horn can be appropriate when you want to signal the car in front of you to GO (the traffic signal turned green) *and* when you want the car to your side to STOP (if it will soon collide with you). What's more, a horn can be motivated by positive emotions (signaling a pedestrian that it's safe to cross in front of you) or negative ones ("Don't even think about cutting in front of me, you sleazebag!").

Laughter, too, can have different meanings in different contexts and be motivated by both positive and negative affective states. Thus, when a friend displays some minor failing, a reminder of our mutual, shared vulnerability is typically born of sympathy and has a reaffirming affect. I label this "Lifting Laughter" (what otherwise may be termed "laughing with" someone). For a similar display by someone who, in our opinion, mistakenly thinks him- or herself our superior, the same message of shared vulnerability can be inspired by feelings of disapproval, what I refer to as "Lowering Laughter" ("laughing at" someone). They are no better than you are. Since individuals also laugh as a result of their own inadequacies (or another's prowess), as well as their own unexpected good fortune, I also describe "Self-Lifting Laughter" and "Self-Lowering Laughter."

One unfortunate side effect of these varied meanings and motivations on the sender's part is the possibility for misinterpretation by receivers. Lifting, Self-Lifting, and Self-Lowering Laughter can sometimes be mistaken for Lowering Laughter.

Because laughter conveys something important about people's feelings, reveals their attitudes about equality, friendship, and status, and makes them feel good when they do it, it seems inevitable that we would at times try to consciously solicit it. These purposeful attempts to generate feelings of amusement are what we typically refer to when we speak of humor. A comprehensive definition of humor would be *any deliberate attempt to inspire a feeling of amusement by creating, manipulating, or highlighting the vulnerability of characters with whom we can identify.* While laughter may be the brass ring, not all attempts at humor attain it. Lesser results are more often the rule than the exception, and we often measure humor on a sliding scale rather than strictly pass/fail.

In the same way that laughter can be motivated by both positive and negative attitudes, humor can be variously described as essentially "Elevating," "Disparaging," "Self-elevating," or "Self-disparaging," although there is no simple correlation between the type of humor and the motivation behind the laughter one hopes to, or will, inspire.

A cardboard belt would be a waist of paper.

The short fortune-teller who escaped from prison was a small medium at large.

I would like to go to Holland one day, wooden shoe?

How do you make antifreeze? Steal her blanket.

What are the two strongest days? Saturday and Sunday. The others are weekdays.

Puns and riddles are amusing when the audience finds them particularly clever. Relative to the pun's creator, the audience member's intellectual standing is slightly diminished, and Self-Lifting Laughter expresses their desire to reaffirm the prior status relationship.

How do you tell whether it's a skunk or a lawyer that's been run over on the highway? There are skid marks in front of the skunk.

Disparaging humor, whether it concerns lawyers, government bureaucrats, or other professions people tend to

be mistrustful of, depends in part on a familiarity with the negative stereotype associated with it. Lawyer jokes are popular in the United States because their ethics obligate them to be advocates for their (often disliked) clients no matter the circumstance. With their loyalty (and passion) appearing to be up for sale, they seem to lack personal traits that most cultures hold in high esteem.

Because most adult Americans are familiar with lawyer jokes, the audience is cued in almost immediately that the question is not intended to be answered seriously and that the joke teller isn't referring to a real-life human having been run over. This sort of emotional "distancing" allows the audience to effectively downgrade what would be, in actuality, a deficiency into a vulnerability. Jokes such as this almost always solicit bit of sympathetic Lifting Laughter from the audience.

What vulnerabilities are highlighted? The most obvious one is physical, since the lawyer wasn't able to avoid losing his life. But perhaps more importantly, we perceive a social vulnerability, in that he was not shown the same consideration drivers grant even to an *animal*. How do we surmise this? If you've ever hit a skunk (or any helpless creature), you know that most people would want to avert it. If you've ever laid down skid marks (or seen someone do it) trying *not* to hit something, you know that's a sign of avoidance. Like most humor, a certain level of knowledge is required to understand who is vulnerable and why.

Who might NOT find humor in this riddle? Those who are exceptionally fond of lawyers; who have no experience with skid marks (the antilock brake generation); who may have had a real friend or family member run over; who are upset with the person telling the joke; who have a humorless lawyer (perhaps their boss) standing next to them…. Many factors affect our appreciation of humor.

Humor With, By, and For AI

Far from being an autonomic, or even subconscious, response to certain incongruities, a feeling of amusement is the product of a highly complex interplay of cognitive, emotional, and social variables. There may be a dozen or more factors that influence its production. In many ways, the process is analogous to an automobile tire's instantaneous loss of traction. Numerous internal characteristics (e.g., rubber formula, width, tread design, air pressure, and wear), as well as external ones (e.g., road surface type, vehicle mass, and weather conditions), affect the timing and degree of skidding, many of which are in place long before the "spontaneous" event occurs.

Similarly, algorithms attempting to evaluate even the simplest forms of humor will require the ability to recognize and measure cognitive elements such as normality, vulnerability, and deficiency as they relate to various human objectives, plus an individual's knowledge base, past experience, and so on. Additionally, they must account for emotional and social features such as the status relationships between various participants, their current moods, the social context, and larger cultural influences.

Even if an AI device was capable of assessing all relevant variables, would it be capable of distinguishing "playful" remarks from serious ones? Except when accurately recounting an amusing, true-life event (which may be considered such only after a sufficient interval of time has passed), humor often relies upon the exaggeration, or outright fabrication, of vulnerabilities. Sometimes it even involves truly horrific things happening, but to completely fictitious (or long dead) characters. Without a full "grasp of reality," could any AI program sufficiently evaluate all these variables?

What if we assumed it could? Many forms of humor ascribe unfavorable qualities (i.e., vulnerabilities) to specific groups based on certain stereotypes, for example, drunken Irishmen, lazy Americans, or stingy Jews. Certain epithets or insults are interpreted by "insiders" as good-natured ribbing or poignant critiques, but by other individuals as racist, sexist, or even blasphemous (e.g., the "N-word" or recent Danish cartoons). Without knowing the device's core beliefs (as if such a thing was possible), would its simulated laugh response be interpreted by human onlookers as expressing tacit agreement with the joke's supposition? Our "sense of humor" says something about us as individuals, our priorities, prejudices, and preferences. The same will be true of anything capable of expressing itself through laughter.

Another aspect of humor made clear by MVT is that its creation may sometimes, but not always, reflect the authentic viewpoints of its creator. We must wonder if AI developers will ever be able to program "sensitivity" to the subtle intricacies of sarcasm. It may be implied (and inferred) by humans with tone-of-voice clues or facial signals, but whether these can be recognized by even the most sophisticated algorithms is questionable. What about playing it safe and refraining from laughter in questionable situations? Unfortunately, in some situations, one's failure to laugh can be just as revealing.

We must also remember that, again, like a car horn, laughter is expressed in varying degrees and for varying durations depending on the level of amusement. It can take the forms ranging from an almost undetectable nasal snort to a brief giggle to a belly laugh and ultimately to rolling-on-the-floor-with-tears laughter. It most cases, laughter's expression is further shaped and moderated by social convention.

Judging the appropriate type and duration of laughter would be yet another important component of any successful algorithm. Indeed, like any other behavior we engage in as humans, laughing can be done "well" and it can be done "poorly." In many instances, someone's ill-timed, inappropriate, or poorly moderated laughter highlights vulnerability on *their* part, and thus acts as a "contagion" for others' laughter (Simon 2008). This possibility, and the potentially negative consequences that might result, must also be considered by AI developers.

In many instances, low levels of amusement don't result in laughter at all, but rather a smile. Smiles can be thought of as expressing low to modest levels of personal vulnerability (Simon 2008). Thus recognizing and using smiles would be an essential tool for a complete range of humor comprehension and response.

If the appreciation of humor by an AI device would be problematic, how much more challenging (and risky) would be its creation? Successful humor always highlights one or more vulnerabilities, either real or imagined. These would need to be ascribed to the device's user, or to a third party, or perhaps to the device itself. Regardless, all involve some risk.

We know from MVT that the difference between elevating humor and disparaging humor, or Lifting Laughter and Lowering Laughter, often depends upon the relationships between the humorist, the target of the humor, and/or the audience—relationships that, in some cases, may have developed over the course of years or even decades. We can only guess the status relationship between user and AI device: friend and confidant, master and servant, or simply operator and instrument.

Something most successful comedians recognize is that outwardly directed humor needs to be balanced with self-deprecating humor. The failure to point out one's own follies as well as those of others can yield a reputation for insensitivity, arrogance, and even bullying. Can this type of self-awareness be programmed into an AI device? More to the point, should it brought to the attention of the user or others and considered a "safer" form of humor?

In the broadest sense of the word, we can think of vulnerability as something inherent to every object or process through which an identifiable objective is sought. Thus any device (perhaps especially one using AI) can be thought of as possessing vulnerabilities. We know that humans have a seemingly innate predilection for anthropomorphism, as evidenced by countless examples of play toys and comedic cat videos. For a device to solicit sympathetic laughter by highlighting its own shortcomings inevitably raises the issue of liability.

Who is responsible for the flaws and limitations exhibited by a device? Would it be the device itself? Not likely. One can assume it will almost certainly be assigned to the user, retailer, manufacturer, or designer. For non-AI tools, certain imperfections are accepted as necessary evils, the result of inevitable compromises when factoring cost, weight, durability, etc. But would the same considerations be afforded a device that is designed to mimic human consciousness *and* actively concedes its own flaws? This is a topic that must be researched well in advance of implementation.

Conclusions

Developing the capacity to correctly distinguish homonyms and homophones will be an important feature of AI devices that respond to verbal instructions. The capacity to recognize humor, however, is more problematic. Even simple forms of humor are difficult to explain using the prevailing conceptual models.

The Mutual Vulnerability Theory provides unique insights into the message conveyed by laughter, its motivations, and the humor designed to solicit it. Multiple variables are assessed as an individual determines whether something or event is amusing. Vulnerability must be highlighted in others or oneself, with the one laughing expressing his/her preference for the status quo ante. This process is entirely dependent upon knowledge base, personal relationships, objectives, personality, mood, and other factors, many in place minutes, months, or years before the specific humor stimulus is encountered.

Humor and the laughter it generates *can* facilitate human interactions and ease anxiety; however, their improper, ill-timed, or malicious use often has the exact opposite effects. AI devices that simulate humor appreciation or attempt its use will be subject to the consequences, positive and negative, realized by their human counterparts. The risk of alienating, insulting, or even provoking a hostile response from others may outweigh the potential benefits. Whatever course the AI community takes, it should advance cautiously using the most comprehensive understanding of laughter and humor available.

References

Fossey, D. 1983. Gorillas in the Mist. Boston: Houghton Mifflin.

Fry, W. F., Jr. 1977. The Appeasement Function of Mirthful Laughter. In A. J. Chapman and H. C. Foot (Eds.), It's a Funny Thing, Humour. Oxford, England: Pergamon.

Morreall, J., ed. 1986. *The Philosophy of Laughter and Humor*. New York: SUNY Press.

Perks, L.G. 2012. The Ancient Roots of Humor Theory. *Humor* 25-2: 119-132.

Provine, R. R. 1993. Laughter Punctuates Speech: Linguistic, Social, and Gender Contexts of Laughter. *Ethology* 95: 291-298.

Simon, J. C. 2008. *Why We Laugh: A New Understanding*. Indianapolis, IN: Starbrook Publishing.

van Lawick-Goodall, J. 1971. In the Shadow of Man. Boston: Houghton Mifflin Company.