1 Introduction

> The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" (I found it!) but "That's funny. . . ."
>
> —Isaac Asimov

Comedy is half of life, according to the theater. (The other half is tragedy.) A large portion, in any case, of people's time is spent attempting to get each other to laugh. Stories are told, jokes recounted, and witticisms cracked whenever possible. In only the most solemn occasions is humor deemed inappropriate, and innovators are pushing the envelope of propriety all the time. When we find humor in a situation, we feel compelled to share it with others. Today, our taste for comedy apparently outstrips our taste for tragedy. Much of our entertainment industry, in every medium (aside from music), consists of humor. If there is not enough comedy in our daily lives, we turn to our televisions and let professional comedians fill the gap, almost in the way we insist on filling our waking hours with recorded music. Like music, alcohol, tobacco, caffeine, and chocolate, humor is a modern human addiction. And if we are to understand humor, we need to adopt a biological perspective from which we can observe—and formulate testable hypotheses about—the evolution of this addiction.

Every cell in our bodies needs sugar—glucose—the fuel that keeps us alive. A good source of glucose is fructose, the sugar in ripe fruit, which the liver can readily convert to glucose. As it turns out, the common natural sugar with the highest subjective sweetness rating—the one the sweetness sensors in our tongues are most tuned to detect—is fructose. So evolution has engineered a powerful fructose-harvesting system and given it a high priority—our cells operate on the rough principle: Whenever the opportunity to harvest fructose is detected, act on it. Honey, which is mainly glucose and fructose, is a particularly good opportunity

for harvesting. It is hard to believe that the yumminess of chocolate cake or maple syrup or strawberry jam all boils down, almost literally, to the deeply practical glucose imperative, but it does. That's the way to understand why we have a sweet tooth. Why do we have a funny bone, a similar craving for, and appreciation of, humor? For a similarly practical reason: We *need* to devote serious time and energy to doing something which, if we didn't do it, would imperil our very lives. . . . Nature has seen to it that we act vigorously on this need, by rewarding that action handsomely.

The phenomena of evolution are not as simple as they are often portrayed. It is not just a matter of the natural selection of "genes for" this or that—whichever feature of living things catches the attention and curiosity of the researcher. In particular, it is important to consider not just the ends but the means, the organic machinery that is going to do the work, whatever it is. The How questions of biology are just as important as the Why questions (Francis 2004), and some evolutionary puzzles are systematically unanswerable without information about the constraints on the performance of the system, and even an educated guess about those constraints depends on having at least a crude model of the machinery. The evolution of our "sense of humor," we will show, could not possibly be explained without hypotheses about the functional architecture of our brains, for the simple reason that what different humorous items have in common is *only* the similar effects they have on those brain systems and the resultant *subjective* experiences. At various points in this book we draw attention to physical complications that really matter, but just as often we slide over complications that we deem—perhaps too riskily—to be ignorable for our purposes. In particular, we set aside for another occasion questions of the complex and dynamical role of development in the relations between genes, organisms, and environment.

As prominent as humor is in our lives, it is at least equally as mysterious. Why does humor exist at all? Why is this category of our experience such a salient feature of our lives? Another question: Why is humor enjoyable? Why shouldn't we simply detect jokes without feeling anything? And why do we laugh (as opposed to belching or scratching our ears, say) when something is funny? These questions are vexing, and our inability to answer them with ease seems at first to be due to our inability so far to answer the question that has led to most of the existing research in humor: What is the *essence* of humor? What features are both necessary and sufficient to differentiate between those things that are funny and those that

are not? We will argue that this question is ill posed; as usual in the post-Darwinian world of biology, it is a mistake to concentrate on finding presumed essential features since one is more likely to find lineages of similar items, evolving according to changing selectional pressures.

The essentialist quandary has two faces. We've just mentioned the difficulty with defining the features for the category of interest, but on the other side there is danger of conflict with nearby categories that may share some of the same features: In the space of human cognitive traits in the neighborhood of humor we also find such categories as nonhumorous riddles, wordplay, and problem solving, as well as other kinds of appreciation of wit and intellect such as the happiness one feels when witnessing a virtuoso performance. Humor experiences blend in with many of these other kinds of experience without clear boundaries between them. Wordplay can be fun without being funny, and so can fishing or gardening or doing one's job. In every case, there can be relatively intense periods where one's emotion borders on glee, and one may even laugh out of sheer pleasure. There is little prospect of drawing a boundary that separates the subspecies *funny* from the genus *delightful*. They are all cognitive joys of one sort or another. Such categories are notoriously difficult to provide with essences (Wittgenstein 1953; Lakoff 1987). We can replace the essentialist question with an improvement: What makes us feel that some things are funny?

This question calls for some sort of causal answer, in terms of processes going on in our minds, and it is our goal to provide a preliminary sketch of not just a cognitive model, but an emotional and *computational* model of humor. This may seem at first to be not just outrageously ambitious, but positively incoherent. The very idea of a *computational* entity that has a sense of humor has long been considered impossible. Even in science-fiction stories that involve artificially intelligent agents (such as the character Data from *Star Trek*), such characters are typically portrayed as lacking the capacity for emotions in general, and especially for particular behaviors such as humor generation and appreciation.[1] The writers of such stories apparently believe that it is not possible to give these traits to a nonbiological computational agent—or else they are tactically conceding this point of ambient prejudice since overcoming it would require too much expository and justificatory effort. We propose to tackle this prejudice

1. But see *Star Trek: The Next Generation*, episode 30: "The Outrageous Okona," in which Data attempts to acquire humor.

head on, arguing that a truly intelligent computational agent could not be engineered *without* humor and some other emotions. These emotions—or their functional equivalents—are requirements of any agent, biological or not, that has human-level intelligence.

When we use the word *computational* here, we intend it more broadly than is typical in cognitive science. We do not yet intend to build a practical testable model, say via neural-network architecture; rather, we are beginning where good design always begins—we want to specify the *functional* requirements of such a computational system so that one day a more technical approach (ideally from computational neuroscience) can provide detailed working blueprints based on the outline we have sketched. We are working toward a theory that would allow humor, as it is experienced—and created—by human beings, to be computed and experienced by a nonhuman agent, a digital machine of some kind that not only can make jokes but that can truly be said to have a "sense of humor" much like the human sense. This is not a straightforward requirement, by any means. At a minimum, it is not sufficient to say that an agent's manifestation of *behavioral expressions* of humor under many or most of the circumstances that elicit such responses in humans indicates a genuine sense of humor in that artificial agent. In order to count as artificial computational humor, the behavioral expression, although necessary as an indicator (how else could it be known that humor was felt?), must also emerge from or be produced by some of the same underlying processing methods and informational contents as natural humor. What aspects of these processes matter? Not the presence of proteins or other biochemical substances, we will argue, but more abstract features of the information-handling processes *and the reasons for their existence*. We will argue that a strict algorithmic approach will be inadequate to imbue an agent with a sense of humor, because the structure of humor is dictated by the riskiness of heuristic processes that have evolved to permit real-time conclusion-leaping, and by the safeguards that have also evolved to protect our minds from these risks. The pivotal causes of genuine amusement and laughter are not simply intrinsic features of the triggering stimuli that are somehow "detected," but *internal* responses that could not be elicited by the triggering stimuli in agents that didn't have a rather specific computational architecture that depends on processes exploited by humorous items.

It will come to light, as we proceed, that computational humor is what we may call an *AI-complete problem*. (In the theory of computation, theorists have developed a classification scheme, in that branch called *complexity theory*, that sorts all computational problems into, roughly, the easy, the

hard, and the "impossible." The most difficult set of problems are called *NP-complete problems*—they require nondeterministic polynomial time to solve, in case you wondered—and if you can solve one of them, you should be able to solve them all.) We use the term *AI-complete* to refer to a class of problems that are no less difficult than the problem of *strong AI* (Searle 1980) or general intelligence—if you can solve any one of them, you've done it by making an artificial agent that *really* thinks.[2] Humor, we will argue, depends on *thought*—it is not just a reflexive response to a stimulus that is inherently funny; it requires a certain category of information processing involving most of the faculties of thought, including memory recall, inference, and semantic integration. It follows, then, that our book must sketch a theory of the kind of general intelligence that could support a genuine sense of humor.

Consider, in contrast, some recent attempts at creating computational humor algorithms. These attempts include JAPE and STANDUP (Binsted 1996; Binsted and Ritchie 2001; Ritchie et al. 2006), WISCRAIC (McKay 2000), and HAHAcronym (Stock and Strapparava 2005). All of these models are algorithmic and syntactical in nature—using punning riddles, phonological word substitution, and acronyms, respectively, as a specific grammatical structure of humorous sentence and then making semantic or phonological substitutions out of lexical tables to create the joke. The largest drawbacks of all the models are that they cannot evaluate the humor they have created, nor can they even be said to know in any sense that they are creating humor. In fact, they do not always create humor; rather, at best they have a higher than chance likelihood of creating a stimulus that can evoke a mildly amused response in humans. They have no critical capacity to understand or evaluate the humor created by others, to say nothing of the capacity to be amused by it.[3] Instead of a "sense of

2. This complexity class has not been proven to have the property of reducibility that is found in complexity theory; take our comparison metaphorically, for now. We're told that the class of AI-complete problems was first described by Fanya Montalvo. Salvatore Attardo may have been the first to apply the similar term "AI-hard" to humor in his book *Humorous Texts: A Semantic and Pragmatic Analysis* (2001).

3. We are aware of one attempt at computational humor detection: Mihalcea and Strapparava (2005) used Naïve Bayes classifiers and Support Vector Machines to separate "one-liner" jokes from other one-line text snippets with impressive results. However, we must interpret these results carefully; these and other machine classification methods notoriously segment the datasets they are given based on features that are not necessarily apparent to the experimenter. In this case, it is very likely that the superficial content or grammatical structure of these one-liners (rather than their effects on the mind) is enough information to suggest which are jokes and

humor," then, they have a very strict generation algorithm reminiscent of traditional grammar-based natural language processing models. Recent research in sentence comprehension suggests that the grammar-based model of language processing does not describe the human mechanism that performs the same job (Jackendoff 2002). We agree, and we will argue moreover that a nonalgorithmic approach is more suited to the problem of comprehension in general, and to the problem of humor comprehension and appreciation in particular.

As we have said, we do not yet offer any running computational models. Instead, we will show what features a good computational theory should contain, and what subproblems we will have to solve on a path to getting to that theory. Its key novelties are a new evolutionary explanation of the origin of humor; an ecologically motivated theory of the emotional component of mirth; and a cognitive theory of humor and laughter (based on insightful earlier theories, but made more precise here) that lays out some of the informational and procedural requirements for a computational substrate that could support artificial humor. The base capacity for humor, the innate[4] "funny bone" that provides the underlying machinery without which humor could not exist, is described for the first time, but it is only part of the story. We also deal with how the base capacity has been extensively exploited by our highly social species. We show how the intentional stance—the involuntarily adopted perspective that "automatically" attempts to attribute beliefs and desires to every complex moving thing we encounter—has allowed humor purveyors and aficionados to extend the reach of their art. Being funny is not just for fun; humor has been exapted as a tool in mate selection and sexual competition, allegiance

---

which are not—no cognitive processing is being performed whatsoever. Although not an instance of humor detection, such a computational humor indicator is interesting because it points to cues that humans (or machines) can use to determine whether they are being told a joke before they find the humor in it, thus giving them a head start in looking for mirth-inducing content.

4. A note on nativism: We are aware that claims of innateness may immediately offend the sensibilities of many developmentally minded researchers. Those readers will certainly, and correctly, note that many factors about our subject of study—or any biological subject of study—will be determined through environmental interaction during ontogeny. If you are one of these readers, we ask you to withhold your judgment for just a moment, while we explain: There is certainly a complex developmental path from pure genetic information to the behavioral characteristic of humor; however, if environmental regularities ensure that this path is taken in all healthy members of the species so that some fundamental aspect of the trait is shared in us all, then *in a useful manner of speaking*, the trait is innate. In this fairly regular environment, the genes specify the trait.

probing, belief extraction, and the building of social capital, for instance. Our theory is an unabashedly eclectic theory, drawing heavily on existing work on humor while providing a novel unifying framework for that work that accounts both for the patterns already discerned by generations of earlier humor theorists and for their failure to find a satisfactorily deep account of the biological mechanisms that account for those patterns.

Humor is a hard problem. Consider how wildly diverse a collection you can make of funny things:

1. Puns and wordplay
2. The rubber-faced antics of Jim Carrey or the deadpan gestures of Charlie Chaplin
3. Caricatures
4. Situation comedies
5. Musical jokes
6. Cartoons
7. "Real-world" humor, the perhaps uncategorizable *objets trouvés* that occur in daily life, and cause us to laugh, whether or not they get turned into items of comedy

What could these possibly have in common—aside from the fact that they can all be very funny? This baffling diversity (and there's more) tempts everyone to concentrate on a few favored genres that work well for one's theory and set aside the others "for the time being." Moreover, everywhere one looks, one discovers the lack of sharp boundaries or thresholds. For instance, some caricatures are entertaining without being amusing, some provoke a smile or a chuckle, and others are downright hilarious; the spectrum of wordplay runs from intriguing puzzles to laugh-provoking puns, with every intermediate shade well exemplified. To make matters worse, there is tremendous variability in who finds what funny. Humor is heavily dependent on shared background assumptions, moods, and attitudes. Then there are the secondary effects or metaeffects, such as the pleasure that a good joke brings to someone who has heard it before, a pleasure that is less "emotional" than "intellectual"—the appreciation from a critical standpoint of the excellence of design of the particular item. (This is like a chef's pleasure in just thinking about the perfect sauce for some dish.)

Taking the evolutionary perspective seriously is the only way, we think, of finding the unity in this diversity. Before Darwin articulated his theory of evolution by natural selection, life forms were bafflingly diverse—what did they have in common aside from being alive? Darwin drew on

a vast repository of excellently observed and codified natural history, a magnificent database waiting to be turned into evidence by a suitably fundamental theory. Following his example, we will canvass the treasury of earlier work on what might be called the natural history of humor, taking advantage of the many insightful analyses and observations to be found there and trying to show how to position them into a theoretical structure that can explain both the patterns and the exceptions.