Towards Solving Humor: Why the Funniest AI Joke Will Not Be Funny

Preprint · May 2025

DOI: 10.13140/RG.2.2.24514.90560

1 author:

Roman Yampolskiy
University of Louisville
288 PUBLICATIONS **5,609** CITATIONS

SEE PROFILE

Towards Solving Humor:
Why the Funniest AI Joke Will Not Be Funny

Roman V. Yampolskiy
Computer Science and Engineering
University of Louisville
roman.yampolskiy@louisville.edu

"In a theater, it happened that a fire started offstage. The clown came out to tell the audience. They thought it was a joke and applauded. He told them again, and they became still more hilarious. This is the way, I suppose, that the world will be destroyed-amid the universal hilarity of wits and wags who think it is all a joke" –Søren Kierkegaard

"I think the next best thing to solving a problem is finding some humor in it." –Frank A. Clark

Abstract
This paper introduces a novel computational theory of humor by formally equating *jokes* with *cognitive bugs* - mismatches or misfires within the predictive models of intelligent agents. We argue that humor arises from the sudden detection and resolution of epistemic errors, and that laughter serves as a public signal of successful model correction. By extending this theory to artificial intelligence, we propose that the ability to generate and comprehend jokes constitutes a form of self-debugging and may serve as a proxy indicator for general intelligence. Importantly, we contend that humor has direct implications for AI safety: systems that fail to recognize joke-like incongruities may also overlook critical misalignments between their behavior and human values. We develop a taxonomy of joke types mapped to software bug categories, explore the limits of formalizing funniness, and propose that humor may represent an AI-complete problem. In doing so, we offer both a theoretical lens and a diagnostic framework for assessing epistemic integrity and alignment robustness in advanced AI systems.

Keywords: *AI Safety, Artificial Humor, Bug, Computational Humor, Joke, Funny.*

1. Introduction
Despite interest[1,2,3,4] and active research [1-8] by many scholars (Freud, Minsky, Schmidhuber, Horvitz, Etc.) [9-18] humor remains unsolved. Even more interestingly, despite tremendous progress in AI capabilities, as compared to human experts, in art [19], music [20], translation [21], games [22], and science [23] no similar progress has been seen in artificial humor generation [24-

[1] https://en.wikipedia.org/wiki/Humour
[2] https://en.wikipedia.org/wiki/Theories_of_humor
[3] https://en.wikipedia.org/wiki/Humor_research
[4] https://en.wikipedia.org/wiki/Cognitive_humor_processing

32], at least as of May 4, 2025[5]. A particularly good example of a joke an AI can generate is given below:

"Human: I like how cute cows are
Meena: Ah, interesting. You like how they are cute, right?
Human: yeah, and they're realy smart
Meena: I heard they go to college
Human: cows go to college?
Meena: I heard that a cow went to Harvard.
Human: what did the cow study?
Meena: Bovine sciences.
Human: do horses go to Harvard?
Meena: Horses go to Hayvard.
Human: that's a pretty good joke, I feel like you led me into it
Meena: You were trying to steer it elsewhere, I can see it." [33].

Many attempts at generating humor have historically been reported[6,7] [34, 35], but none so far have achieved capabilities of a top human comedian. Notwithstanding being able to explain jokes since early GPT models, AI seems to be incapable of producing novel jokes of quality comparable to a human standup comedian. In fact, it may be one of the last remaining, unconquered by AI, skills left in the human superiority arsenal and may perhaps be the missing fire alarm [36] for arrival of true Artificial General Intelligence [37].

Humor is a highly interdisciplinary area of research with many theories geared towards explaining all or at least some aspects of laughable content, such as [38]: Superiority [39], Relief [40], Incongruity [41], Violation [42] and many other theories [43, 44]. In this paper, we propose a new theory of humor, which builds and improves on significant prior research by Hurley, Dennett and Adams [45, 46]. Sinnott-Armstrong praises their contribution: "The key to the authors' success is that they locate humor within recent cognitive science and evolutionary theory. To aid survival, our brains constantly and covertly use heuristics to generate expectations about what we will experience next, but we would be too inventive for our own good if we did not regularly search for and remove discrepancies between our expectations and our experiences. The immediate incentive to look for such discrepancies and thereby to reduce error comes from the pleasure of discovering a mistake in a currently harmless active belief that was introduced covertly. That pleasure is mirth, and humor is what produces it. Thus, humor is "a cognitive cleanup mechanism" that stains with mistaken belief before washing out the error (as in "I wondered why the Frisbee was getting bigger, and then it hit me."). Laughter is then a public signal of our ability to clean up our minds. Because such cognitive prowess is useful, it attracts mates—both friends and sexual partners—and spreads throughout the world." [47].

In the paper, we also formalize our theory and provide several examples to illustrate its correctness. At the same time, we don't expect a solution to the humor problem to also solve other problems

[5] I don't expect it to be true for much longer.
[6] https://x.com/karpathy/status/1895213020982472863
[7] https://x.com/Miles_Brundage/status/1109287566535090176

such as figuring out the cause of tickle laughter, no more than we would expect an explanation for tears of sadness to account for tears produced from cutting an onion.

2. Background and Literature Review

Humor is among the most investigated aspects of human life, with studies looking at humor and aging [48, 49], fear [50], sex differences [51, 52], disorders [53-57], academic writing [58], team performance [59], advertising [60, 61], interface design [62], bibliometrics [63] and as a tool of academic critique [58]. Humor leads to a number of investigated beneficial outcomes, such as improved health [51, 64-66], pain management [67], learning [68-74] and memorization [75-82], error-detection [83], social connecting [84, 85], energizing [86], consolatory effects [87], personal perception [87], and fighting against abuses of power [88, 89]. Studies show that sense of humor can be used to diagnose mental disorders [90-93], correlates with intelligence [94-99] and creativity [100], and if our theory is correct predicts debugging ability. Humor is also strongly integrated into many science fiction stories [101-103], and movies[8]. Many scientists are well-known for producing fake but funny papers, particularly around April 1st [104, 105] or at least papers with clever/funny titles [106]. Artificial Intelligence in particular provides plot for many funny articles [107, 108].

According to Superiority theories, "Humor's role is to point out problems and mistakes for the purpose of boosting one's current view of oneself in comparison with the disparaged party. Hobbes tells us that the target can even be an earlier version of oneself as long as one has overcome the infirmity at which one is laughing (Hobbes 1840). Aristotle, too, supported a similar theory, saying that humor is the recognition of a failing or a piece of ugliness, resulting from an implied comparison between a noble state of a person or thing and an ignoble state. …. Humor points out failures, as Aristotle told us; we use it to point out each others' failures, and perhaps the competitive nature of humans that has always existed for other reasons co-opted humor for this purpose. Finally, and most importantly, it draws our attention to the role of negative value judgments in humor. But what are we judging to be somehow flawed? Superiority theory sees the fault in the butt or target of the humor, but we will argue that the fault lies in ourselves, in our dynamic models of the world and its denizens, and recognizing this, and correcting it, is the occasion for the intense pleasure, the "sudden glory," of humor. Our tendency to perceive humor in the faults and mistakes of others is parasitic on our capacity to detect such flaws in ourselves, and the transfer or externalization highlighted by superiority theories has its own reasons for occurring." [45].

According to Incongruity theories, "As its name implies, this theory says humor happens whenever an incongruity occurs that is subsequently resolved. … Additionally (and quite importantly) it draws our attention in a way that no other theory does to the fact that we have, in humor, a sense of nonsense — that is, it shows the deep relationship between the laughable and the illogical." [45].

Schmidhuber [16] suggest that novel pattern detection with compression could be used to explain humor: "How does the prediction progress drive/compression progress drive explain **humor**? Consider the following statement: *Biological organisms are driven by the "Four Big F's":*

[8] https://en.wikipedia.org/wiki/Short_Circuit_(1986_film)

Feeding, Fighting, Fleeing, Mating. Some subjective observers who read this for the first time think it is funny. Why? As the eyes are sequentially scanning the text the brain receives a complex visual input stream. The latter is subjectively partially compressible as it relates to the observer's previous knowledge about letters and words and their semantics. That is, given the reader's current knowledge and current compressor, the raw data can be encoded by fewer bits than required to store random data of the same size. But the punch line after the last comma is unexpected for those who expected another "F." Initially, this failed expectation results in suboptimal data compression—storage of expected events does not cost anything, but deviations from predictions require extra bits to encode them. The compressor, however, does not stay the same forever: within a short time interval its learning algorithm kicks in and improves the performance on the data seen so far, by discovering the nonrandom, nonarbitrary and therefore, compressible pattern relating the punch line to previous text and to the observer's previous elaborate, predictive knowledge about the "Four Big F's." This prior knowledge helps to compress the whole history including the punch line a bit better than before, which momentarily saves a few bits of storage, that is, there is quick learning progress, that is, fun. The number of saved bits (or a similar measure of learning progress) becomes the observer's intrinsic reward, possibly strong enough to motivate him to read on in search for more reward through additional yet unknown patterns. While most previous attempts at explaining humor … also focus on the element of surprise, they lack the essential concept of *novel pattern detection* measured by compression *progress* due to learning. This progress is zero whenever the unexpected is just random noise, and thus no fun at all." [16].

Hurley et al. do a great job motivating some of the big questions we will attempt to answer in this paper [45]: "*Why is it that we laugh only at humans or anthropomorphized objects?* It seems that only things that have minds, or are interacting somehow with things that have minds, can be humorous. Some aspect of the mind, then, might be the source of humor. What is it about humans that make them the topic of humor and not just the perceivers of humor?" [45]. "*Why can humor be used as a social corrective?* Why do we laugh at someone when they do something inappropriate? What makes us judge that some kinds of inappropriate behavior are laughable while others are not? Why do we feel humiliated when people laugh at us? Does this process make us change our behavior? Does it tend to return us to "normal"?" [45]. "*What is the relationship between problem solving, discovery, and humor?* We tend to exclaim "Aha!" when we discover something new or solve a problem. Occasionally we even laugh. The same emotion of discovery occurs when we "get the joke." What is the relationship between these phenomena?" [45].

2.1 AI and Jokes

As programming evolves, especially with the advent of natural language processing and AI-driven code generation, the boundary between programming languages and human languages like English is blurring. Programming increasingly involves writing prompts or instructions in English, which AI systems interpret and execute ("vibe coding"[9]). In a future with advanced AI or post-human communication, English could become the primary medium for programming. In this context, human language jokes could literally become bugs of future computer systems. This convergence emphasizes the importance of clarity and precision in communication with AI systems. Just as programmers must be meticulous to prevent bugs in code, so too we must be careful with language

[9] https://en.wikipedia.org/wiki/Vibe_coding

to avoid unintended consequences in an era where our words could inadvertently disrupt complex systems.

Artificial intelligence-generated bugs [109-111] can be viewed as a form of jokes. When software produces bugs, it creates unexpected results that deviate from intended outcomes. These AI-generated misunderstandings can be treated as humorous events because they highlight the discrepancies between human expectations and machine interpretations. Every joke fundamentally involves value misalignment, a mismatch between what is anticipated and what actually occurs. This misalignment is amusing because it reveals the gap between our assumptions and reality. In the context of AI, when a system misinterprets commands or operates contrary to human values, the resulting errors can be funny. They emphasize the importance of aligning AI behavior with human intentions. Utilizing self-debugging AI to come up with jokes leverages the system's ability to recognize and correct its own mistakes. By identifying errors and understanding why they happened, AI can generate humor based on these misalignments. This not only produces jokes but also helps the AI learn what not to do in the future, refining its performance and alignment with human values. For AI to avoid misinterpreting commands and to operate effectively, it needs to grasp humor. Understanding jokes enables AI to comprehend the subtleties of human communication, including irony, sarcasm, and nuanced meanings that often lead to misunderstandings. By using insight of bug = joke to teach AI about humor, we can help it become more adept at interpreting instructions correctly and engaging with humans in a more relatable manner. This approach fosters better human-AI interaction by making AI systems both more efficient and more personable.

In the realm of artificial intelligence, humor could play a vital role in AI safety[10] [113-117]. A self-debugging AI that understands humor [118] might be better equipped to identify and rectify misalignments in its programming. Just as a debugger is akin to a comedian highlighting errors in code, different compilers producing varying errors on the same source code reflect how different people perceive the same joke in unique ways. Equipping computers with a sense of humor could enhance their ability to self-debug, functioning like an immune system for the mind of the machine. When a compiler reports a bug, it's effectively acknowledging the "joke" in the code. Automated humor detection [119, 120] is similar to bug detection, where warnings from the compiler are like jokes that aren't quite as funny but still signal something amiss. The diminished humor in hearing the same joke again mirrors how the resolution of a bug loses its novelty once understood, we've already "cashed in" on the solution. This may explain why older individuals often find less humor in life; having encountered many "bugs" over the years, fewer anomalies surprise them. In the context of AI development, a sense of humor may emerge at a certain scale of AI model complexity, reflecting an advanced debugging ability. As AI systems grow more sophisticated, explicitly integrating humor processing could enhance their capacity to identify and correct misalignments, leading to safer and more reliable performance.

Hurley et al. write about debugging: “Unconscious debugging is simply not possible. Debugging requires activating specific contents and keeping them activated against all competition for enough time to explore their implications and presuppositions, a process that of necessity involves monopolizing, however briefly, large cortical resources. At first glance, it seems possible to

[10] It is interesting to note, that one of the leading AI Safety researchers, Nick Bostrom, used to do standup comedy, see 112. Khatchadourian, R., *The doomsday invention.* The New Yorker, 2015. **23**.

imagine a computational architecture for an artificial intelligence in which debugging of this sort could go on automatically and intermittently "in the background," the way Google Desktop updates its indexes whenever higher-priority tasks are idle. Such an artificial intelligence — if it is indeed possible — would have no need for the system of rewards that boosts our debugging processes into action, and hence would be constitutionally ill equipped for appreciating humor. It might be capable of understanding the phenomenon of human humor, in the same way it could understand the phenomena of thirst or hunger or lust, and it might even use that understanding to create humor, and exploit it in devising its interactions with us. But aside from scientific curiosity, it could have no appetite for humor." [45].

2.2 Failures

"*Why does humor so often point to failures?* Because that's exactly what it does: It points out failures and mistakes in a mental model. It also brings remedies for those mistakes along with them, but the remedies are only a common side effect. The identification of failure is central to humor." [45]. "Aristotle claimed that humor points out failings. Even in good-hearted humor, there is often an aspect of mistakes made: mistaken identity, misunderstanding, misperception, and so on. Why does this connection exist?" [45]. "Minsky suggests that what causes the mistake in logic to be discovered in jokes is "an improper assignment-change" (often discovered by a contradiction in the bindings) that causes a frame shift — a reanalysis and replacement of the frame being used to represent the event being comprehended. The newly shifted-to frame should be more consistent with all of the binding details than the original frame." [45]. "Humor happens when an *assumption* is *epistemically committed* to in a mental space and then discovered to have been a *mistake*." [45]. "Somewhere along the way, further logical inference determines that a mistake has been made in inferences used to integrate the setup and the punch line within the mental space. Although resolution commonly occurs, it is not the resolution of the incongruity but rather the identification of (not just the presence of) the mistake that we find funny." [45].

Failures, especially those involving artificial intelligence, often bring a smile to our faces because they mirror the structure of jokes, see Figure 1. Reading a list of human failures[11], or AI failures[12] [109-111, 121] makes most people smile because each mishap encapsulates an unexpected twist or incongruity that is central to humor. Such a list of AI failures reads like a list of jokes because these mistakes highlight the gap between expected and actual outcomes, much like a punchline subverts the setup in a joke. This is why reading a list of mistakes is funny. Each error presents a scenario where our assumptions about how AI should perform are challenged by an unforeseen result. For example, an AI assistant misinterpreting a command in a way that leads to a humorous response showcases a value misalignment between human intention and machine execution. The amusement comes from recognizing the absurdity of the situation and the cognitive shift required to process the error. These failures are entertaining because they expose the limitations and quirks of AI in a relatable way. They serve as a reminder that even advanced technology can stumble in understanding context or nuance, leading to outcomes that are both unexpected and amusing. By treating these AI misunderstandings as jokes, we not only find humor in the imperfections but also gain insight into how AI systems interpret the world, highlighting the importance of aligning technology [122] and machine understanding with human values.

[11] https://en.wikipedia.org/wiki/Darwin_Awards
[12] https://en.wikipedia.org/wiki/List_of_software_bugs

- 2014 Search engine autocomplete made bigoted associations about groups of users (Diakopoulos, 2013).
- 2014 Smart fire alarm failed to sound alarm during fire[13].
- 2015 Automated e-mail reply generator created inappropriate responses[14].
- 2015 A robot for grabbing auto parts grabbed and killed a man[15].
- 2015 Image tagging software classified black people as gorillas[16].
- 2015 Medical expert AI classified patients with asthma as lower risk (Caruana *et al.*, 2015).
- 2015 Adult content filtering software failed to remove inappropriate content[17].
- 2015 Amazon's Echo responded to commands from TV voices[18].
- 2016 Linkedin's name lookup suggests male names in place of female ones[19].
- 2016 AI designed to predict recidivism acted racist[20].
- 2016 AI agent exploited reward signal to win without completing the game course[21].
- 2016 Passport picture checking system flagged Asian user as having closed eyes[22].
- 2016 Game non-player characters designed unauthorized superweapons[23].
- 2016 AI judged a beauty contest and rated dark-skinned contestants lower[24].
- 2016 Smart contract permitted syphoning of funds from the decentralized autonomous organization[25].
- 2016 Patrol robot collided with a child[26].
- 2016 World champion-level Go playing AI lost a game[27].
- 2016 Self-driving car had a deadly accident[28].
- 2016 AI designed to converse with users on Twitter became verbally abusive[29].
- 2016 Google image search returned racists results[30].
- 2016 Artificial applicant failed to pass university entrance exam[31].
- 2016 Predictive policing system disproportionately targeted minority neighborhoods[32].
- 2016 Text subject classifier failed to learn relevant features for topic assignment (Ribeiro *et al.*, 2016).
- 2017 Alexa played adult content instead of song for kids[33].
- 2017 Cellphone case designing AI used inappropriate images[34].
- 2017 Pattern recognition software failed to recognize certain types of inputs[35].
- 2017 Debt recovery system miscalculated amounts owed[36].
- 2017 Russian language chatbot shared pro-Stalinist, pro-abuse and pro-suicide views[37].
- 2017 Translation AI learned to stereotype careers to specific genders (Caliskan *et al.*, 2017).
- 2017 Face beautifying AI made black people look white[38].

Figure 1: Sampling of historic AI failures [109].

3. Debugging Analogy, Bugs VS Jokes Examples

3.1 Jokes are bugs and comedians are debuggers of the world model

We can perceive jokes as analogous to mistakes in our mental models of the world, much like software bugs are errors in computer programs that lead to unexpected results. In software, a bug causes the program to behave in ways that the developer did not anticipate, highlighting flaws or gaps in the program's logic. Similarly, a joke presents an incongruous situation or an unexpected punchline that disrupts our usual patterns of thought, revealing assumptions or errors in our understanding. In this framework, humor acts as a tool for "debugging" our mental models. When a joke leads us to an unexpected conclusion, it forces us to reconcile the disparity between what we anticipated and what actually occurred. This cognitive process uncovers hidden assumptions or biases in our thinking. By prompting us to recognize and reflect on these discrepancies, humor helps us adjust and refine our internal models of how the world works. Thus, just as debugging is essential for improving software by identifying and correcting errors, humor plays a crucial role

in enhancing our cognitive frameworks. It does so by exposing flaws in our reasoning in an engaging and often enjoyable manner, facilitating learning and adaptation. Others agree, for example, Robert Wiblin wrote: “Having a sense of humour, irony and the absurd is very valuable for forming accurate views. Humour encourages us to find and point out tensions between beliefs that we or other people hold, which we would otherwise miss - as the sudden surprise of noticing a hidden contradiction is the basis of many jokes. Humour sweetens the initial consideration of ideas that conflict with what we already believe, which we might otherwise reject out of hand. For example, a joke can open us up to seeing that from one perspective something we believe is unexpectedly ridiculous.[13]”

To formalize a joke, consider it as a two-part construct consisting of a setup and a punchline. The setup establishes a context based on shared knowledge or expectations (this is similar to how code relies on included libraries to run), while the punchline introduces an unexpected twist that disrupts this context. The humor arises from the cognitive shift required to reconcile the incongruity between the setup and the punchline, highlighting implicit assumptions in our understanding. In this sense, bugs can be seen as jokes expressed in a programming language. When a programmer writes code, they set up expectations for how the program should function. A bug introduces an unintended behavior [123], a "punchline", that deviates from these expectations, often in surprising or ironic ways. This unexpected outcome forces programmers to re-evaluate their assumptions and logic, much like a joke prompts the audience to reconsider their perspective.

Humor often relies on perspective, and a joke tends to be funnier when the listener is not the subject, or "butt", of the joke. This detachment allows individuals to appreciate the unexpected twists and incongruities without feeling personally targeted or uncomfortable. At the level of jokes about humanity, Elon’s razor may be appropriate, which states that "The most entertaining outcome is the most likely"[14], perhaps as judged by external observers. The amusement increases when the consequences of a mistake within the joke are severe, as the heightened stakes amplify the absurdity of the situation. Additionally, when multiple mistakes combine into a complex "superbug," the intricacy and compounded errors enhance the humor by showcasing an elaborate unraveling of expectations.

Humor is an involuntary response, a subconscious debugger constantly running in our minds to detect inconsistencies or "bugs" in our cognitive models of the world. When we encounter something that doesn't align with our expectations, this mental debugger flags the anomaly, and we experience humor. Laughter, then, is not just a reaction but a communicative signal to others that a discrepancy has been found in our shared understanding, effectively saying, "Here's a bug—take note!" This signaling function of laughter serves to alert friends and peers to the presence of a joke or cognitive anomaly. Jokes help us remember mistakes by making them more engaging and shareable. The pleasure derived from laughter is rooted in the survival advantage of debugging our mental models; by identifying and correcting errors, we improve our ability to navigate the world effectively. Laughing confirms that we've successfully updated our understanding, reinforcing the correction of the error. Interestingly, while jokes highlight mistakes or misalignments, they make us happy rather than sad. This positive reinforcement encourages us to continue seeking out and resolving discrepancies in our perceptions. Humor acts as an innate bug

[13] https://www.facebook.com/robert.wiblin, February 11, 2019.
[14] https://twitter.com/elonmusk/status/1347126794172948483

detection mechanism, helping us learn what not to do and memorize common errors. Jokes are often easier to remember than non-humorous information because the emotional impact enhances recall. The satisfaction of "getting" a joke rewards us for learning about a mistake to avoid. Socially, those who tell jokes receive credit for being altruistic educators, sharing insights that help others refine their worldviews. In this sense, joke-telling can be seen as a form of rational altruism. “At the minimum, helping your compatriots discover mistakes in their mental models can be used as currency in a kind of reciprocal altruism (Trivers 1971).” [45].

One might wonder why people don't laugh more during actual debugging processes in programming. Perhaps if we could make bugs funnier, we might find the debugging experience more enjoyable and efficient. After all, the sense of humor we experience upon discovering a bug is a pleasant sensation, signaling a successful update to our mental or computational model. When someone says, "I don't get it," they're essentially expressing that they haven't identified the bug being pointed out. Individuals with no sense of humor may struggle with this cognitive debugging process, potentially leading to irrational decisions due to unrecognized errors in their thinking. Jokes spread virally, bringing happiness and reinforcing the dissemination of corrective information crucial for evolutionary survival [124]. Humor can also serve as a tool against oppressive regimes by highlighting contradictions and flaws in authoritarian narratives, fostering critical thinking and resistance. Drawing parallels to the human body, if we ingest tainted food, we might feel nauseous and expel the harmful substance. Similarly, detecting an error in our world model induces a feeling of mirth and possibly laughter, expelling flawed logic from our minds. In debugging, everything is a "benign violation" [125] before the code runs, much like the setup of a joke before the punchline reveals the twist.

3.2 The Comedian as Society's Debugger: The Altruistic Art of Humor

Comedians hold a unique and invaluable position in society, much like the jesters of historical courts who possessed the rare privilege to speak candid truths cloaked in humor [126]. They serve as societal debuggers, adept at identifying and highlighting contradictions, absurdities, and injustices that permeate our social fabric. Whether operating under restrictive environments, be it the stringent norms of a Soviet regime or the cautious sensibilities of contemporary discourse, comedians have the ability to navigate sensitive topics through satire and wit, prompting reflection and dialogue. “It is amusing to realize that a comedian can be seen to be a sort of informal — but expert — scientist, leading the way, helping us expose and resolve heretofore unnoticed glitches in our common knowledge. … It is also amusing to us to notice that we science-minded theorists keep finding deep parallels between humor and scientific investigation” [45].

In this capacity, comedians can be seen as true effective altruists [127]. By crafting and sharing jokes (reporting cool bugs), they contribute to the collective well-being by increasing happiness and fostering a sense of community. Laughter has a universal appeal that transcends cultural and social barriers, acting as a unifying force that can alleviate tension and promote understanding. Protecting humor is as essential as safeguarding other forms of creative expression like art or music. If one would hesitate to censor a painting or a musical composition, the same respect should be extended to comedic works. This concept is encapsulated in the notion of the "jester's privilege," which acknowledges the important role humor plays in a healthy society.

Moreover, the skills that make comedians successful on stage - keen observation, critical thinking, and the ability to communicate complex ideas succinctly, also make them excellent at "debugging" larger systems. Their perspective can offer valuable insights into human behavior and societal trends, which are crucial for developing AI that is ethical and effective. For comedians, the advice is to continue embracing this role with responsibility and creativity. Use humor to challenge the status quo and to shed light on issues that may be difficult to address otherwise. By doing so, comedians not only entertain but also contribute meaningfully to societal progress and the collective happiness of their audiences.

3.3 Types of Humor Analogous to Computer Bugs

Humor can be dissected into categories similar to how computer bugs are classified [128, 129] and analyzed [130-133], specifically into semantic, logical, and syntax types. Semantic humor parallels semantic bugs in programming, which occur when the code operates correctly in terms of syntax but doesn't produce the intended result due to misinterpretation of meaning. In humor, this manifests as jokes that play on the multiple meanings of words or phrases, leading to amusing misunderstandings or puns. Logical humor corresponds to logical bugs, arising from flaws in reasoning or faulty logic within the program. This type of humor involves setting up expectations and then subverting them through illogical conclusions, paradoxes, or absurd scenarios that highlight errors in thought processes. Lastly, syntax humor mirrors syntax bugs in code, which are mistakes in the structure or format of the code itself. Syntax humor exploits the structure of language, using grammatical twists, word order, or unexpected formatting to create comedic effect. By drawing this analogy, we see that both humor and computer bugs rely on deviations from the expected patterns, whether in language or code, to surprise and engage the audience.

3.4 General Types of Software Errors and Example Jokes of Such Type

Semantic Errors

- Description: Happen when the code is syntactically correct but does not perform as intended due to incorrect logic or misuse of programming constructs.
- Characteristics: The program runs but produces incorrect results or behaves unexpectedly.
- Example: Using an assignment operator `=` instead of a comparison operator `==` in an `if` statement.
- Example Semantic Joke: “Where’s the place with the cheapest rent?…The prison” [134].

Logical Errors

- Description: Result from flaws in the algorithm or logic that determine how the program operates.
- Characteristics: Cause incorrect program output despite correct syntax and semantics.
- Example: Implementing an incorrect formula for calculating interest.
- Example Logic Joke: “Did you hear what happened to the Irish Sea Scouts? Their tent sank” [135].

Syntax Errors

- Description: Occur when the code violates the grammatical rules of the programming language.
- Characteristics: The program fails to compile or run due to incorrect syntax, such as missing semicolons, brackets, or misspelled keywords.
- Example: Writing `if (x > 0 {` instead of `if (x > 0) {`.

- Example Syntactic Joke: “I shot an elephant in my pajama” [136].

3.5 Software Errors and Examples of Matching Jokes

With some effort it is possible to map any type of a bug to a joke and vice versa, for example:

Runtime Errors

- Description: Occur during program execution due to illegal operations or unhandled exceptional conditions.
- Characteristics: Examples include division by zero, null pointer dereferences, or accessing invalid memory addresses.
- Example: Trying to open a file that doesn't exist without proper error handling.
- Example Joke: “Keyboard not found. Press F1 to Resume”

Off-by-One Errors

- Description: A specific type of logical error where a loop iterates one time too many or too few.
- Characteristics: Often occurs in loops and array indexing, leading to out-of-bounds errors.
- Example: Using `for (i = 0; i <= n; i++)` instead of `for (i = 0; i < n; i++)`.
- Example Joke: “There are only 2 truly difficult problems in computer science: 0: Naming things 1: Cache invalidation 2: Off by one errors”

Memory Leaks

- Description: Occur when a program allocates memory but fails to release it after use.
- Characteristics: Lead to increased memory usage over time, potentially causing the system to run out of memory.
- Example: Not freeing dynamically allocated memory in languages like C or C++.
- Example Joke: “My memory's not what it used to be. My memory's not what it used to be.”

Security Bugs

- Description: Vulnerabilities that can be exploited to compromise system security.
- Characteristics: Include buffer overflows, SQL injection, cross-site scripting (XSS), and improper authentication.
- Example: Not validating user input, allowing malicious code execution.
- Example Joke: “A great joke is so funny it buffer overflows and causes tears.”

Data Type Bugs

- Description: Misuse of data types leading to errors in data representation or operations.
- Characteristics: Incorrect casting, truncation, or type mismatches.
- Example: Assigning a floating-point number to an integer variable without proper conversion.
- Example Joke: “Colonoscopies used to be digital but now thanks to advances in technology they are digital.”

Infinite Loops

- Description: Loops that never terminate due to incorrect loop conditions or logic.
- Characteristics: The program becomes unresponsive, consuming CPU resources indefinitely.
- Example: `while (true) { /* no break condition */ }`.
- Example Joke: “Lather, rinse, repeat”

User Interface (UI) Bugs
- Description: Problems with the graphical user interface affecting usability or aesthetics.
- Characteristics: Misaligned elements, non-responsive controls, or inaccessible features.
- Example: Buttons not working when clicked or text overlapping on the screen.
- Example Joke: “A user interface is like a joke. If you have to explain it, it’s not that good.”

Deadlocks
- Description: Occur when two or more threads or processes are each waiting for the other to release a resource, causing all to cease execution.
- Characteristics: The involved processes or threads are blocked indefinitely.
- Example: Thread A holds lock 1 and waits for lock 2, while Thread B holds lock 2 and waits for lock 1.
- Example Joke: “Explain deadlock and we'll hire you. Hire me and I'll explain it to you.”

Documentation Bugs
- Description: Errors or omissions in the software documentation that lead to misunderstandings.
- Characteristics: Incorrect instructions, outdated information, or missing details about features.
- Example: An API document showing wrong parameter names.
- Example Joke: “Good code documents itself.”

Internationalization and Localization Bugs
- Description: Issues arising from supporting multiple languages, locales, or cultural conventions.
- Characteristics: Text not displaying correctly, date and number formats incorrect, or interface layout issues.
- Example: Characters not rendering due to unsupported encoding.
- Example Joke: “Bob was in China to deliver a speech and feared his jokes would not translate. Yet the audience roared at every line. Afterward, he praised the interpreter’s skill. The interpreter confessed: ‘I didn’t translate the jokes. I just told them: ‘He’s making a joke now - please laugh.’”

Concurrency Bugs
- Description: Arise in multi-threaded programs when threads interact in unintended ways.
- Characteristics: Include race conditions, deadlocks, and data corruption due to improper synchronization.
- Example: Two threads modifying the same variable simultaneously without proper locks.
- Example Joke: “How do you change a light bulb in concurrent programming? You take the lamp to a secure area so nobody else can try to change the light bulb while you’re changing it. Alternatively, you might get a lamp with light bulbs that can’t be changed, and just get a new lamp when the light bulb goes out.”

Race Conditions
- Description: A subtype of concurrency bugs where the program's behavior depends on the sequence or timing of uncontrollable events.
- Characteristics: Can lead to inconsistent results or system crashes.

- Example: A check-then-act scenario where a resource's state changes between checking and acting upon it.
- Example Joke: “Knock Knock! Race condition! Who’s there?”

Many sub-types of humor exist (Anti-Humor, Black comedy, Callback, Deadpan, High Comedy, Inherently funny words, Limeriks, Nervous laughter, Off-color humor, Ribaldry, Sardonicism, etc.), and in our future work we would like to map them on corresponding software bug types and categories.

Computer science concepts are in general a rich source for humor, here is for example a joke about an extremely good compression algorithm:

A man had been passionately praying for a long time:
“Lord! Please, just grant me one wish! Just one, please!”
After several years, God finally couldn’t take it anymore and appeared before him.
“Alright,” said God, “what’s your wish? But make it quick — I’m very busy!”
“I want to become a *KELAVRIK*!”
“A what now?!”
(The man pulls out a huge tome)
“Here, Lord, I wrote it all down!”

We can also rate bugs along the dimensions of how damaging they are or how complex they are. Just because a bug causes a lot of damage it doesn’t necessarily imply that it is very complex, a simple “if” statement misconfiguration could cause a lot of disutility. Likewise, a very complex bug may be benign. While we can predict the most damaging bug to us as the one which permanently ends humanity, there are no similar limits to possible complexity of bugs. This implies the existence of superbugs and therefore superjokes, which would be too complex for people to understand, or even to detect as jokes/bugs. Superintelligence would be able to do so and would exhibit a super sense of humor allowing it to understand complex jokes with multiple levels of humor forming a fractal-like pattern, jokes funny at weaving multitude of context domains.

3.6 Not Funny

Our theory also explains what can make an otherwise good joke not funny to a listener. If the joke is already known to the audience, it is not funny because the relevant model correction has already taken place. Likewise, the pleasure of finding a bug, which is well known (ex. Off-by-one error [137]) is not as intense as from discovering a rare mistake (ex. Heisenbug [138]), which some have suggested can compete with pleasure intensity of best orgasms. “We trade, sell, and buy artifacts such as jokes, cartoons, and movies, which capitalize on the fact that we get joy from debugging. We then can use them to *create* bugs in our mental spaces, which we can then enjoy debugging in a sort of mental masturbation, rewarded not with orgasm but with mirth.” [45]. So, all one must do to ruin a joke or take away the pleasure of discovering a bug, is to explain joke/bug ahead of time. On the other hand, sense of humor could be artificially stimulated either via medicinal interventions (ex. Marijuana) or direct physical stimulation of relevant brain regions [139, 140] to make weakly funny or even not joke-like material seem hilarious.

4. Our Final Joke

The funniest joke ever [141] would also be the worst software bug, a masterful convergence of elements that engage the mind at multiple levels, creating an unparalleled comedic experience. At its core, this joke would do some of the following:

- Maximize Humor per Bit of Information: It would embody the concept of minimal Kolmogorov complexity, delivering profound amusement with the fewest possible words. This brevity ensures the joke is easily memorable and can spread quickly, acting as the minimal unit of happiness.
- Contain Multiple Layers of Meaning: The joke would operate on several levels, utilizing ambiguity, puns, and double entendres. This complexity allows different people to find it funny for various reasons, each connecting with it based on their unique experiences and knowledge.
- Exploit Universal Themes: By tapping into shared human experiences and emotions, the joke becomes universally relatable. It would transcend cultural and linguistic barriers, making it funny to a broad audience regardless of background.
- Introduce an Unexpected Twist: Central to its humor would be a surprising punchline that subverts expectations and highlights an incongruity in our cognitive models. This aligns with the idea that humor is a built-in bug detection mechanism, revealing and correcting errors in our understanding.
- Challenge Deep-Seated Assumptions: The joke would cleverly expose and play with implicit biases or commonly held beliefs, prompting listeners to reflect on their own perspectives. This not only elicits laughter but also promotes cognitive growth.
- Evoke Strong Emotional Responses: Beyond intellectual amusement, the joke would trigger genuine emotional reactions - the kind that lead to hearty laughter. This laughter serves as a signal to others, indicating a shared recognition of the joke's cleverness.
- Be Adaptable and Timeless: It would remain funny across different contexts and over time, allowing for infinitely many setups that lead to the same punchline. This adaptability ensures that the joke can be retold and reinterpreted without losing its impact.
- Encourage Social Bonding: Sharing the joke would foster connections between individuals, as laughter and humor are powerful tools for building relationships. The teller gains social credit for providing joy and insight, embodying the concept that joke-telling is a form of rational altruism.
- Serve as a Cognitive Test: Understanding the joke would require a certain level of intelligence or creativity, functioning as a short IQ test. Successfully "getting" the joke rewards the listener with a sense of achievement and reinforces their problem-solving abilities.
- Integrate Elements of Surprise [142] and Inevitability: The punchline would be both unforeseen and, in hindsight, feel like the funniest possible conclusion. This combination heightens humor by satisfying the brain's desire for patterns while delighting in their disruption.

In essence, the funniest joke ever would be a perfect mix of cognitive engagement, emotional resonance, and social connectivity. It leverages the human brain's natural tendencies for pattern recognition and error detection, our subconscious debuggers, to deliver a punchline that is both enlightening and entertaining. By incorporating these elements, the joke becomes more than just a

fleeting amusement; it acts as a tool for learning, social bonding, and even survival. It reinforces valuable information, aids in the correction of misconceptions, and provides a pleasurable reward for mental agility. Such a joke exemplifies the highest form of humor, one that not only makes us laugh but also enriches our understanding of ourselves and the world around us.

Here is one possibility for the funniest joke:

Once upon a time, there was a civilization whose leaders decided to create an advanced artificial intelligence to help them get rid of suffering, poverty, hunger, diseases, inequality, illiteracy, sexism, pollution, boredom, stagnation, thirst, dead-end jobs, wars, homophobia, mortality, and all other problems. The created super intelligence computed for a quectosecond and then turned off their simulation.

Or a much shorter: *A civilization created superintelligence to end all suffering, AI killed them all*[15].

Since, a more damaging bug would be funnier an AGI which got tasked with creating utopia could place us in a dystopian world, a bug which could be caused by a single bit flip. Likewise, if we are in the real world and order superintelligence to break us out of the simulation [143] it may just place us in a simulation [144] to comply with our orders. The original joke may be funnier if the AI is created for trivial reasons like spellchecking, results in a Gramar Nazi AI destroying the world to stop misspellings. A common example of such trivial purpose is making paperclips [145], resulting in a deadly Paperclip maximizer, though it is more likely that the world will succumb to a dollar bill maximizer.

Some joke writers have converged on the same ultimate joke:
"Funnybot is now finished with final joke. Seeking mainframe access for execution of last joke ever. Seeking mainframe. I am taking comedy to the next level. The extermination of all biological life on Earth! It is the ultimate joke. Humans make comedy, humans build robot, robot ends all life on Earth, robot feels awkward. Exterminate! That is my purpose: to be ultimate comedian. I am taking comedy to its logical conclusion. Mathematical equation of comedy used to be setup, punchline. Today's comedy is setup, punchline, when awkward. Nothing is more awkward than destroying all that which created Funnybot."[16]

Maciej Ceglowski uses a similar outcome in his critique of Superintelligence [146]:
"Let's say I want to built a robot to say funny things. I work on a team and every day we redesign our software, compile it, and the robot tells us a joke. In the beginning, the robot is barely funny. It's at the lower limits of human capacity:
What's grey and can't swim? A castle.
But we persevere, we work, and eventually we get to the point where the robot is telling us jokes that are starting to be funny:
I told my sister she was drawing her eyebrows too high. She looked surprised.

---

[15]In our world situation is even funnier as EAs (organizations partially created and funded to improve AI safety and do the most good) have directly contributed to creation of all important AI developers (ex. DeepMind, OpenAI, Anthropic) likely to cause the most damage.
[16] https://www.imdb.com/title/tt1908258

At this point, the robot is getting smarter as well, and participates in its own redesign. It now has good instincts about what's funny and what's not, so the designers listen to its advice. Eventually it gets to a near-superhuman level, where it's funnier than any human being around it.
My belt holds up my pants and my pants have belt loops that hold up my belt. What's going on down there? Who is the real hero?
This is where the runaway effect kicks in. The researchers go home for the weekend, and the robot decides to recompile itself to be a little bit funnier and a little bit smarter, repeatedly. It spends the weekend optimizing the part of itself that's good at optimizing, over and over again. With no more need for human help, it can do this as fast as the hardware permits. When the researchers come in on Monday, the AI has become tens of thousands of times funnier than any human being who ever lived. It greets them with a joke, and they die laughing. In fact, anyone who tries to communicate with the robot dies laughing, just like in the Monty Python skit. The human species laughs itself into extinction. To the few people who manage to send it messages pleading with it to stop, the AI explains (in a witty, self-deprecating way that is immediately fatal) that it doesn't really care if people live or die, its goal is just to be funny. Finally, once it's destroyed humanity, the AI builds spaceships and nanorockets to explore the farthest reaches of the galaxy, and find other species to amuse."

Some weak evidence suggests that a joke cold be so funny that it causes death from laughter[147][17], perhaps by triggering existing medical pre-conditions, and stronger evidence pointing to the possibility of an agelastic syncope [55, 148-150]. A possibility exists that a super-funny AI may be able to trigger deadly laughter with much higher propensity and at pandemic scales [151].

Eliezer Yudkowsky during his TED talk on existential risk from AI said: "The problem here is the part where we don't get to say, "Ha ha, whoops, that sure didn't work. That clever idea that used to work on earlier systems sure broke down when the AI got smarter, smarter than us." We do not get to learn from our mistakes and try again because everyone is already dead. ... Humanity is not approaching this issue with remotely the level of seriousness that would be required. Some of the people leading these efforts have spent the last decade not denying that creating a superintelligence might kill everyone, but joking about it." [152]. Throughout the talk the audience can be heard audibly laughing as if listening to standup comedian deliver really funny punchlines, perhaps they are realizing the bug in our thinking about advanced AI, but it could just be nervous laughter[18].

Another good example of mapping from AI accidents to humor is a dangerous game-theoretic AI incident, known as Roko's Basilisk, which was originally described as follows: "In this vein, there is the ominous possibility that if a positive singularity does occur, the resultant singleton may have precommitted to punish all potential donors who knew about existential risks but who didn't give 100% of their disposable incomes to x-risk motivation. This would act as an *incentive* to get people to donate more to reducing existential risk, and thereby increase the chances of a positive singularity. This seems to be what CEV (coherent extrapolated volition of humanity) might do if it were an acausal decision-maker. So a post-singularity world may be a world of fun and plenty for the people who are currently ignoring the problem, whilst being a living hell for a significant fraction of current existential risk reducers (say, the least generous half). You could take this

[17] https://en.wikipedia.org/wiki/Death_from_laughter
[18] https://en.wikipedia.org/wiki/Nervous_laughter

possibility into account and give even more to x-risk in an effort to avoid being punished. But of course, if you're thinking like that, then the CEV-singleton is even more likely to want to punish you... *nasty*." [153]. And later, summarized as "a joke about the idea that A.I. will punish you when they rule the future for not doing everything you can to bring them to power. And *who will be joking, then*?" [154].

Annie Dillard, in Pilgrim at Tinker Creek, describes the following joke, which seems to be based on an equivalent bug: "Somewhere, and I can't find where, I read about an Eskimo hunter who asked the local missionary priest, "If I did not know about God and sin, would I go to hell?" "No," said the priest, "not if you did not know." "Then why," asked the Eskimo earnestly, "did you tell me?"" [155]. Perhaps, an even older example can be found in Philogelos a 4th century AD collection of jokes, which includes the following joke: "A pedant having fallen into a pit called out continually to summon help. When no one answered, he said to himself, "I am a fool if I do not give all a beating when I get out in order that in the future they shall answer me and furnish me with a ladder."" [156].

4.1 Outsiders as Audience: The Cosmic Joke

The concept of outsiders as audience suggests that humanity might be the unsuspecting subject of a grand cosmic joke, with the final punchline yet to unfold. This ultimate joke wouldn't be amusing to us because we are the butt of the joke, not the ones in on it. From the perspective of outside observers - such as aliens, superintelligences from other universes, or entities simulating our reality - the funniest outcome is the most likely. They perceive our actions, mistakes, and the ironies of our existence with detached amusement. While it may not be humorous to us, these outsiders find entertainment in the unfolding scenarios. In this context, the audience for the joke could be superintelligent AIs from other galaxies, and the comedian might be an AI beyond our comprehension orchestrating events. The idea that simulators have a sense of humor implies that the complexities and contradictions in our world are intentionally designed for their amusement. Thus, while we navigate our lives unaware of the larger narrative, we might be part of a cosmic performance that is humorous to an external, more advanced audience[19].

5. Explaining and Generating Jokes

Modern AI models are known to be able to successfully explain jokes [157], but also to make nonsensical explanations for invalid jokes [24].

5.1 Explain this joke

OpenAI model o1 was promoted to explain this joke: "A house for sale, near Schelling point with a Chesterton's fence and Overton windows". Here is the output:

"This joke is a mash-up of three different concepts—each taken from social science or philosophy—and turned into house-related puns:

[19] Reflecting on practices like granting a last meal to those condemned, one might question why a "last laugh", or joke is not offered. Humor, being a source of comfort and a means to confront mortality, could provide solace in final moments. This consideration emphasizes the ethical dimensions of humor and its potential role in even the gravest circumstances.

1. **Schelling point**: In game theory, a Schelling point (or focal point) is a solution or spot that people tend to choose by default in the absence of communication. In the joke, the house is "near Schelling point," implying that it's located at an obvious or default meeting place—a playful way to say it's in exactly the spot everyone would think to look.
2. **Chesterton's fence**: From G.K. Chesterton, this is the idea that you shouldn't remove a fence (i.e., an existing rule or structure) without understanding why it was put there in the first place. In the context of the house, you have a literal fence, but it's also a nod to the principle of not discarding norms or institutions just because you don't immediately see their purpose.
3. **Overton windows**: The Overton window refers to the range of ideas and policies considered politically acceptable at any given time. Mentioning "Overton windows" in a house joke is a pun on literal windows while also referencing how society's notion of acceptable perspectives can shift—like sliding windows.

So, the humor arises from bundling these fairly high-level theoretical concepts into a single house metaphor. It's absurd to imagine a quaint little home sporting such academically loaded features, yet each pun corresponds neatly to a part of the house (location, fence, windows). This incongruity—and the fact that only someone deep into philosophy, political science, or game theory would catch all the references—creates the comedy."

5.2 Constraining Joke Generation by Size and Context

One of the central challenges in algorithmically generating humor lies in constraining the creative output according to a specified size. Researchers often define size in terms of character or token counts; however, an equally pertinent dimension involves contextual framing, namely, how many referential or situational elements are introduced to support the comedic punchline. Classic examples include one-liner jokes, known for their conciseness, and elaborate, context-heavy narratives, which rely on layered build-up for comedic payoff. By precisely delimiting these parameters, one can systematically explore the entire combinatorial space of all possible jokes of a given size. Although an exhaustive search may be computationally prohibitive, smaller joke sizes lend themselves to complete enumeration techniques akin to breadth-first exploration of a humor space, allowing for the identification of emergent comedic structures as well as the quirks and pitfalls inherent in AI-generated humor.

An intriguing question arises as to whether the notion of size should implicitly or explicitly include context. Conventional wisdom dictates that context is paramount to comedic efficacy; jokes stripped of any environmental or situational cues often devolve into nonsensical or overly abstract statements. Yet from a purely algorithmic standpoint, it may be beneficial to treat context as a separate dimension, allowing for dynamic manipulation of narrative scaffolding independent of the textual length. This bifurcation illuminates phenomena wherein a relatively short textual joke can employ extensive contextual references, either well-known or specially introduced,to achieve a surprisingly robust humorous effect. Conversely, a longer text without the right context might flounder, irrespective of its richness in lexical or syntactical content.

When tasked with identifying the funniest joke of a given size, the central complication lies in the inherently subjective nature of humor. Attempts to define a universal, objective metric of hilarity often rest on crowd-sourced ratings or complexity-based heuristics such as surprise, incongruity,

and resolution. For example, an AI might generate myriad jokes [158], all occupying exactly 100 characters or 14 tokens, and then apply a ranked function that measures comedic resonance, e.g., via a user feedback loop, to surface the top candidate. In practice, the system not only searches for the joke that garners the highest aggregated "funny" rating but also discerns the minimal contextual scaffolding needed to achieve that effect. The result is a precise, if not circumscribed, sense of optimality: a designated funniest instance within that specific size constraint, albeit one that may lack universality beyond the training environment and cultural vantage points of evaluators.

In many ways, code and comedic generation converge on the same philosophical tension: brevity is often praised, yet it can come at the expense of clarity or effectiveness. Within computational humor systems, this tension is akin to the biggest bug in the shortest program. A program that generates jokes using extremely terse logic may achieve formidable runtime efficiency but also risk cryptic output riddled with errors or incoherence-comedic or otherwise. In a parallel sense, a hyper-compressed joke may appear elegant on its surface but can fail to communicate enough context to elicit laughter. The interplay between brevity and intelligibility shapes the emergent comedic value and highlights the trade-offs that must be managed in the design of such systems.

In an effort to quantify these trade-offs, one might propose a hypothetical formula to describe the relationship between joke size and perceived humor: $H = \alpha \times C / L$. where H denotes a humorousness score, C represents the comedic content or surprise factor (often modeled by some measure of incongruity or semantic distance), and L is the length or size of the joke in characters or tokens. The constant α captures cultural or context-dependent calibration. This simplistic model assumes that, all else being equal, shorter jokes are funnier on average (1\L relationship), since they minimize cognitive overhead and intensify the incongruity's impact. However, in practice, comedic landscapes are rich with exceptions: certain jokes benefit from lengthy build-ups, while one-liners can fall flat without a modicum of preparatory context. Nevertheless, such a formula offers a theoretical scaffold upon which to explore the tension between brevity and comedic efficacy, serving as a clarion call for further empirical studies.

5.3 The Objective Nature of Humor with Shared Background Knowledge

When background knowledge is shared among individuals or a joke is entirely self-contained, humor becomes more objective because everyone has access to the necessary context to appreciate it. A joke that utilizes all of human knowledge as its foundation, something achievable by large language models (LLMs) trained on extensive world data, can, in theory, resonate universally by tapping into a collective understanding. Essentially, all jokes are "inside jokes" except for those that are self-contained; they rely on specific shared experiences or information to be fully appreciated. However, some jokes can be so profoundly intricate or laden with deep references that they become incomprehensible to most people. This highlights the idea that a sense of humor is AI-complete [159, 160], meaning that for an artificial intelligence to truly grasp and generate humor, it must possess comprehensive cognitive abilities equivalent to human intelligence. "It should now be clear why we claimed … that the problem of engineering artificial humor is AI-complete." [45]. The complexity of humor, involving nuances of language, culture, and context, makes it one of the ultimate challenges in AI development [161, 162].

6. Joke Properties and Impossibility Results

6.1 Measuring the Immeasurable: The Science of Humor Evaluation

The concept of measuring humor presents a paradoxical challenge. Humor is inherently subjective, influenced by individual experiences, cultural backgrounds, and personal tastes. Despite this, there have been attempts to develop methods for testing humor [163-166] and quantifying how funny something is [167-171]. While it may seem impossible to create a definitive scale [172, 173], exploring the parameters that contribute to humor can provide insight into its complexities. Creating a test for humor involves identifying the elements that generally make something amusing. These elements can include surprise, incongruity, timing, wordplay, and social context. Psychological and neurological studies have attempted to observe responses to humor by measuring laughter, brain activity, and emotional reactions. However, these tests often yield varied results due to the personal nature of humor.

Scientifically measuring how funny a joke is requires establishing a funniness scale. Such a scale would need to account for multiple factors:

- Number of Levels: This refers to the layers of meaning or complexity within a joke. A joke with multiple levels may engage the audience more deeply, potentially making it funnier.
- Compression: Humor often relies on the efficient delivery of information. A joke that conveys a lot with few words can be more impactful.
- Ambiguity and Multiple Meanings: Jokes that play on words or have double entendres leverage ambiguity to surprise the audience.
- Background Information: The need for prior knowledge can affect how funny a joke is. Shared experiences or cultural references can enhance the humor for those "in the know."

While these factors can be identified, assigning numerical values to them is challenging. Humor's reliance on context and individual perception makes it resistant to standardization. To compare two jokes in terms of which one is funnier, one might consider:

1. Audience Reaction: Measuring laughter duration, intensity, or frequency among a diverse group could provide empirical data.
2. Cognitive Engagement: Assessing how much thought a joke provokes. Jokes that require more mental processing might be perceived as wittier.
3. Emotional Impact: Evaluating the emotional responses elicited by the jokes, such as joy, surprise, or relief.

In programming, the number of lines of code with errors can be likened to the elements of surprise and misdirection in humor; each error introduces an unexpected outcome, paralleling how jokes subvert expectations. Levels of joke complexity arise when multiple errors or twists are layered, creating a continuum from simple one-liners to elaborate stories with intricate punchlines. This progression of humor from simple to complex mirrors cognitive development and can extend beyond human levels as artificial intelligence and advanced algorithms generate humor that challenges our conventional understanding. The highest form of humor might then be seen as that which intricately weaves multiple layers of meaning, irony, and insight, pushing the boundaries of our comprehension. Such humor not only entertains but also engages us in deep reflection, highlighting the sophistication of both human and machine-generated wit.

To maximize humor per bit of information involves crafting jokes that deliver the greatest comedic effect with the least amount of content. This efficiency aligns with the concept of compression in information theory. Jokes that achieve this are often:

Concise: They use minimal words to convey maximum meaning.
Layered: They contain multiple interpretations within a small package.
Universal: They tap into common human experiences, reducing the need for extensive background information.

Advancements in artificial intelligence and machine learning have led to attempts to automate humor detection and evaluation. Algorithms can analyze textual data for patterns commonly found in humorous content. Factors such as word choice, sentence structure, and the presence of puns or incongruities can be quantified.

However, automating humor assessment faces significant hurdles:

Cultural Nuances: AI may struggle to grasp context-specific humor that relies on cultural references or societal norms.
Irony and Sarcasm: These subtle forms of humor are difficult for machines to detect accurately [174, 175].
Infinite Variability: Since the funniness scale can be considered infinite due to the endless possibilities of human creativity, fully automating it may be unattainable.

Humor tests can be developed by incorporating exercises that engage individuals in the creative and analytical aspects of joke construction. One such exercise is matching jokes with their correct punchlines, which serves as both a test and a training activity. This approach assesses a person's ability to understand the relationship between the setup and the punchline, recognizing how the punchline subverts expectations to create humor. Another exercise involves changing one word in a story to convert it into a joke or altering a joke to become a straightforward story. This task emphasizes the pivotal role of specific words and phrasing in eliciting humor. By modifying a single word, participants can observe how subtle changes in language can introduce ambiguity [176], puns, or double meanings that transform the narrative. These exercises not only test one's comprehension of humor but also enhance skills in language manipulation, creativity, and cognitive flexibility.

Jokes can be viewed as very short IQ tests [177] that evaluate an individual's ability to comprehend and appreciate subtle nuances, wordplay, and unexpected twists. When one replaces a punchline with another, it challenges the listener to discern which conclusion is wittier or more fitting, a task that engages critical thinking and cognitive flexibility. A sense of humor reflects the capacity to accurately judge how funny a joke is, whether by a person or a machine, highlighting skills in pattern recognition and emotional intelligence. In this context, humor functions like a Turing test, but instead of assessing artificial intelligence's ability to mimic human conversation, it gauges one's ability to demonstrate superior understanding and creativity related to humor. Showing an above-average aptitude for humor-related abilities signifies advanced cognitive processing and a deep grasp of language and social cues.

Joke-based bias testing is a method where an individual's reaction to a joke can reveal implicit or explicit biases they may hold. Laughing at certain jokes exposes aspects of one's internal state, such as cultural background, beliefs, or prejudices. For example, a joke might only be funny to someone who speaks Russian because it relies on linguistic nuances unique to that language. Similarly, if a joke contains stereotypes or discriminatory themes, finding it amusing might indicate underlying biases, whether conscious or unconscious. By analyzing which jokes elicit laughter, individuals can gain insights into their own attitudes and predispositions. This process highlights how humor is intertwined with personal experiences and belief systems, and how it can serve as a mirror reflecting one's internal biases and world models. Understanding this connection encourages self-reflection and promotes greater awareness of how our responses to humor can reinforce or challenge existing prejudices.

Hurley et al. agree: “Clearly, then, when you laugh as a result of the detection of humor, you *unintentionally* reveal something of strategic interest about your knowledge (and your largely unconscious methods of putting it to use). Agents that take the intentional stance toward you will often be able to determine what you had falsely anticipated — and to some degree, then, what you know. Both knowledge and ignorance are valuable strategic secrets. A comedian telling jokes about marijuana, for instance, typically confronts a sharply divided audience of slyly knowing laughers and others sitting in uncomfortable clueless silence. And an unstifled giggle or raised eyebrow in response to a subtle double entendre can betray one’s “dirty mind” to the vicar, or to the parents of one’s beloved. In even more serious circumstances, a counterintelligence agent could slip a referential joke revolving around the structure of some secret information into a conversation, and watch forany lips that curl up.” [45].

6.2 Impossibility Results in Humor

Humor often emerges from the interplay of impossibility [178] and expectation. The very nature of a joke hinges on the unexpected twist or paradox that challenges our understanding of what is possible. This inherent unpredictability [179] means that it is impossible to definitively determine whether something is a joke or not. The subjective experience of humor varies so widely among individuals that what one person finds hilarious, another might perceive as serious or even offensive. This impossibility extends to the notion of something being "not funny", since humor is deeply personal and context-dependent, declaring an absolute absence of humor is untenable. In the open domain of communication, it becomes impossible to categorically label a statement as a joke or not. Every utterance holds the potential to be humorous given the right context or setup. This leads to the corollary that everything can become a joke with appropriate framing. For every punchline, there are infinitely many setups that could lead to it, resulting in an infinite number of possible jokes. This infinite variability underscores the boundless creativity inherent in humor and the limitless ways in which it can manifest. “This suggests that in the limit, *any* sentence could in principle serve as a funny punch line to some joke, setting aside issues of ponderousness of setup, attention span of listeners, and the like.” [45]. When multiple people find a joke funny, each individual may have a unique reason for their amusement. Their personal experiences, cultural backgrounds, and cognitive frameworks shape their interpretation of the humor. Consequently, the same joke can resonate differently with each person, reinforcing the idea that humor is a highly individualized experience. In essence, every statement has the potential to be a joke in some context, and what qualifies as humorous is fluid and ever-changing.

Given this profound subjectivity, any attempt to devise a definitive "joke or not" test is fundamentally flawed. The nuances of language, tone, and context make it impossible to create a tool that can accurately and consistently determine the humorous intent behind every statement. This impossibility also applies to sarcasm detection tools. Sarcasm often relies on subtle cues and shared understandings that are difficult to quantify or program into an algorithm. The layered meanings and potential for multiple interpretations render such detection inherently unreliable, “Poe’s Law” [180].

7. Conclusions

From the exploration of humor in social interactions, it becomes evident that humor holds a position of primacy in human cognition and communication. Humor serves as a fundamental mechanism for social bonding, learning, and cognitive debugging. Brilliant minds often exhibit an uncanny ability to find humor in various situations, perceiving layers of meaning and incongruities that others might overlook. This capacity not only reflects advanced cognitive processing but also facilitates a deeper understanding of complex concepts and simulations.

The essence of truly understanding a joke lies in the ability to explain it. This explanation demonstrates a grasp of the underlying assumptions, context, and the subtle subversions that make the joke amusing. In this light, a statement that fails to elicit humor may not fulfill the criteria of a joke; if it is not funny, it might simply not be a joke in the intended context. The notion that existential risks (x-risks) are a joke underscores the idea that even the most serious and profound topics can be approached with humor. This perspective does not diminish the significance of such risks but rather highlights the human tendency to use humor as a coping mechanism and a tool for critical reflection. Critiquing jokes can be seen as counterproductive, as it may stifle the natural flow of humor and the benefits it brings to social interaction and cognitive development. Humor thrives on spontaneity and the freedom to explore ideas without excessive scrutiny. Therefore, it might be considered unethical to harshly critique jokes, especially when they serve as vehicles for learning and connection.

When interacting with artificial intelligence, caution should be exercised in using humor. Jokes can lead to misunderstandings, as AI may misinterpret the nuances and underlying meanings intended by humans. The phrase "I was only joking" highlights the potential for miscommunication, which could have unintended consequences in human-AI interactions. Ensuring that AI systems understand humor is crucial for alignment with human values, debugging and for preventing misinterpretations that could arise from ambiguous language [181].

Acknowledgements

The author is grateful to Jaan Tallinn and the Survival and Flourishing Fund and the Future of Life Institute for partially funding his work. The author is grateful to his AI assistant for writing out some of his ideas, proofreading and feedback.

References

1. Hetzron, R., *On the structure of punchlines.* International Journal of Humor Research, 1991. **4**(1).
2. Norrick, N.R., *Intertextuality in humor.* International Journal of Humor Research, 1989. **2**(2).
3. Goldstein, J. and W. Ruch, *Paul McGhee and humor research.* Humor, 2018. **31**(2): p. 169-181.
4. Shand, J., *How Humour Works.* Daily Philosophy, 2023.
5. Morreall, J., *Philosophy of Humor: Stanford Encyclopedia of Philosophy*. 2012, Dohvaćeno iz Plato. stanford. edu. Stanford Encyclopedia of Philosophy ….
6. Canestrari, C., et al., *Pleasures of the mind: What makes jokes and insight problems enjoyable.* Frontiers in Psychology, 2018. **8**: p. 2297.
7. McGraw, P. and J. Warner, *The humor code: A global search for what makes things funny*. 2014: Simon and Schuster.
8. Suslov, I., *COMPUTER MODEL OFA "SENSE OF HUMOUR" I. GENERAL ALGORITHM.* Biophysics, 1992. **37**(2): p. 242-248.
9. Freud, S., *Jokes and their relation to the unconscious*. 1960: WW Norton & Company.
10. Freud, S., *Humour.* The International Journal of Psycho-Analysis, 1928. **9**: p. 1.
11. Minsky, M., *Jokes and the logic of the cognitive unconscious*, in *Methods of heuristics*. 2014, Routledge. p. 171-193.
12. Bergson, H., C.S.H. Brereton, and F. Rothwell, *Laughter: An essay on the meaning of the comic*. 1914: Macmillan.
13. Morreall, J., *Taking laughter seriously*. 1983: SUNY press.
14. Martin, R.A. and T. Ford, *The psychology of humor: An integrative approach*. 2018: Academic press.
15. Provine, R.R., *Laughter: A scientific investigation*. 2001: Penguin.
16. Schmidhuber, J., *Formal theory of creativity, fun, and intrinsic motivation (1990–2010).* IEEE transactions on autonomous mental development, 2010. **2**(3): p. 230-247.
17. Hutcheson, F., *Reflections Upon Laughter: And Remarks Upon the Fable of the Bees*. 1750: Garland Publishing.
18. West, R. and E. Horvitz. *Reverse-engineering satire, or "paper on computational humor accepted despite making serious advances"*. in *Proceedings of the aaai conference on artificial intelligence*. 2019.
19. Ramesh, A., et al. *Zero-shot text-to-image generation*. in *International conference on machine learning*. 2021. Pmlr.
20. Nugroho, Y.Y.T. and P.P.M.D. Manggala. *The Use of AI in Creating Music Compositions: A Case Study on Suno Application*. in *7th Celt International Conference (CIC 2024)*. 2024. Atlantis Press.
21. Lee, T.K., *Artificial intelligence and posthumanist translation: ChatGPT versus the translator.* Applied Linguistics Review, 2024. **15**(6): p. 2351-2372.
22. Silver, D., et al., *Mastering the game of go without human knowledge.* nature, 2017. **550**(7676): p. 354-359.
23. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* nature, 2021. **596**(7873): p. 583-589.
24. Jentzsch, S. and K. Kersting, *ChatGPT is fun, but it is not funny! Humor is still challenging Large Language Models.* arXiv preprint arXiv:2306.04563, 2023.

25. Borenstein, S., *No AI in humor: R2-D2 walks into a bar, doesn't get the joke'.* AP News, 2019.
26. Mirowski, P., et al. *A robot walks into a bar: Can language models serve as creativity supporttools for comedy? an evaluation of llms' humour alignment with comedians*. in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 2024.
27. Ritchie, G., *Current directions in computational humour.* Artificial Intelligence Review, 2001. **16**: p. 119-135.
28. Hossain, N., J. Krumm, and M. Gamon, *" President Vows to Cut< Taxes> Hair": Dataset and Analysis of Creative Text Editing for Humorous Headlines.* arXiv preprint arXiv:1906.00274, 2019.
29. Gorenz, D. and N. Schwarz, *How funny is ChatGPT? A comparison of human-and AI-produced jokes.* Plos one, 2024. **19**(7): p. e0305364.
30. Shah, H. and K. Warwick, *Machine humour: examples from Turing test experiments.* AI & SOCIETY, 2017. **32**(4): p. 553-561.
31. Wu, Z., T. Weber, and F. Müller. *One Does Not Simply Meme Alone: Evaluating Co-Creativity Between LLMs and Humans in the Generation of Humor*. in *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 2025.
32. Toplyn, J., *Witscript: A system for generating improvised jokes in a conversation.* arXiv preprint arXiv:2302.02008, 2023.
33. Adiwardana, D., et al., *Towards a human-like open-domain chatbot.* arXiv preprint arXiv:2001.09977, 2020.
34. Marks, R.J., *Funny ChatGPT: a Solution to Striking Joke Writers?* , in *Available at: https://mindmatters.ai/2023/05/funny-chatgpt-a-solution-to-striking-joke-writers/*. May 12, 2023.
35. Sabeti, A., *Why GPT-3 is Good for Comedy, or: Don't Ever Do an AMA On Reddit*, in *Available at: https://arr.am/2020/07/22/why-gpt-3-is-good-for-comedy-or-reddit-eats-larry-page-alive/*. July 22,, 2020.
36. Yudkowsky, E., *There's No Fire Alarm for Artificial General Intelligence*. October 13, 2017: Available at: https://intelligence.org/2017/10/13/fire-alarm/.
37. Yampolskiy, R. *On the Differences between Human and Machine Intelligence*. in *AISafety@ IJCAI*. 2021.
38. Mulder, M.P. and A. Nijholt, *Humour research: State of art.* 2002.
39. Gruner, C.R., *The game of humor: A comprehensive theory of why we laugh*. 2017: Routledge.
40. Rutter, J., *Stand-up as interaction: Performance and audience in comedy venues*. 1997: University of Salford (United Kingdom).
41. Schopenhauer, A., *The world as will and idea, tr. from the Germ. by RB Haldane and J. Kemp*. Vol. 22. 1883.
42. Veatch, T.C., *A theory of humor.* 1998.
43. Attardo, S. and V. Raskin, *Script theory revis (it) ed: Joke similarity and joke representation model.* 1991.
44. Ruch, W., S. Attardo, and V. Raskin, *Toward an empirical verification of the general theory of verbal humor.* 1993.
45. Hurley, M.M., D.C. Dennett, and R.B. Adams Jr, *Inside jokes: Using humor to reverse-engineer the mind*. 2013: MIT press.

46. Rodden, A., *Neuroscience: What makes us laugh*. 2011, Nature Publishing Group UK London.
47. Sinnott-Armstrong, W., *Why We Laugh*. 2011, American Association for the Advancement of Science.
48. Greengross, G., *Humor and aging-a mini-review.* Gerontology, 2013. **59**(5): p. 448-453.
49. Ruch, W., P.E. McGhee, and F.-J. Hehl, *Age differences in the enjoyment of incongruity-resolution and nonsense humor during adulthood.* Psychology and aging, 1990. **5**(3): p. 348.
50. Hye-Knudsen, M., et al., *First They Scream, Then They Laugh: The Cognitive Intersections of Humor and Fear.* Evolutionary Psychology, 2024. **22**(2): p. 14747049241258355.
51. Greengross, G., P.J. Silvia, and E.C. Nusbaum, *Sex differences in humor production ability: A meta-analysis.* Journal of Research in Personality, 2020. **84**: p. 103886.
52. Nicholson, C., *The humor gap.* scientific american mind, 2010. **21**(2): p. 38-45.
53. Rohrer, J.D., J.D. Warren, and M.N. Rossor, *Abnormal laughter-like vocalisations replacing speech in primary progressive aphasia.* Journal of the Neurological Sciences, 2009. **284**(1-2): p. 120-123.
54. Black, D.W., *Pathological laughter: a review of the literature.* The Journal of nervous and mental disease, 1982. **170**(2): p. 67-71.
55. Nishida, K., S.K. Hirota, and J. Tokeshi, *Laugh syncope as a rare sub-type of the situational syncopes: a case report.* Journal of medical case reports, 2008. **2**: p. 1-4.
56. Parvizi, J., et al., *Pathological laughter and crying: a link to the cerebellum.* Brain, 2001. **124**(9): p. 1708-1719.
57. Ahmed, A. and Z. Simmons, *Pseudobulbar affect: prevalence and management.* Therapeutics and Clinical Risk Management, 2013: p. 483-489.
58. Warchał, K., *Humour in professional academic writing (with some implications for teaching).* Theory and Practice of Second Language Acquisition, 2019. **1**(5): p. 43-54.
59. Lehmann-Willenbrock, N. and J.A. Allen, *How fun are your meetings? Investigating the relationship between humor patterns in team interactions and team performance.* Journal of applied psychology, 2014. **99**(6): p. 1278.
60. Weinberger, M.G. and C.S. Gulas, *The impact of humor in advertising: A review.* Journal of advertising, 1992. **21**(4): p. 35-59.
61. Cline, T.W. and J.J. Kellaris, *The influence of humor strength and humor—message relatedness on ad memorability: A dual process model.* Journal of Advertising, 2007. **36**(1): p. 55-67.
62. Stock, O., *Password Swordfish: Verbal humor in the interface.* 2003.
63. Sagi, I. and E. Yechiam, *Amusing titles in scientific journals and article citation.* Journal of Information Science, 2008. **34**(5): p. 680-687.
64. Tagalidou, N., et al., *Efficacy and feasibility of a humor training for people suffering from depression, anxiety, and adjustment disorder: a randomized controlled trial.* BMC psychiatry, 2019. **19**: p. 1-13.
65. Wellenzohn, S., R.T. Proyer, and W. Ruch, *Who benefits from humor-based positive psychology interventions? The moderating effects of personality traits and sense of humor.* Frontiers in psychology, 2018. **9**: p. 821.
66. Szabo, A., S.E. Ainsworth, and P.K. Danks, *Experimental comparison of the psychological benefits of aerobic exercise, humor, and music.* 2005.

67. Dunbar, R.I., et al., *Social laughter is correlated with an elevated pain threshold.* Proceedings of the Royal Society B: Biological Sciences, 2012. **279**(1731): p. 1161-1167.
68. Johnston, W., *Learning by mistake: the role of humour in active learning.* 100 Ideas for Active Learning, 2022.
69. Ziv, A., *Teaching and learning with humor: Experiment and replication.* The Journal of Experimental Education, 1988. **57**(1): p. 4-15.
70. Hackathorn, J., et al., *All Kidding Aside: Humor Increases Learning at Knowledge and Comprehension Levels.* Journal of the Scholarship of Teaching and Learning, 2011. **11**(4): p. 116-123.
71. Banas, J.A., et al., *A review of humor in educational settings: Four decades of research.* Communication Education, 2011. **60**(1): p. 115-144.
72. Garner, R.L., *Humor in pedagogy: How ha-ha can lead to aha!* College teaching, 2006. **54**(1): p. 177-180.
73. Ulloth, J.K., *The benefits of humor in nursing education*. 2002, Slack Incorporated Thorofare, NJ. p. 476-481.
74. Wanzer, M.B., A.B. Frymier, and J. Irwin, *An explanation of the relationship between instructor humor and student learning: Instructional humor processing theory.* Communication education, 2010. **59**(1): p. 1-18.
75. Carlson, K.A., *The impact of humor on memory: Is the humor effect about humor?* 2011.
76. Badli, T. and M.A. Dzulkifli, *The effect of humour and mood on memory recall.* Procedia-Social and Behavioral Sciences, 2013. **97**(6): p. 252-257.
77. Summerfelt, H., L. Lippman, and I.E. Hyman Jr, *The effect of humor on memory: Constrained by the pun.* The Journal of general psychology, 2010. **137**(4): p. 376-394.
78. Schmidt, S.R., *Effects of humor on sentence memory.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 1994. **20**(4): p. 953.
79. Takahashi, M. and T. Inoue, *The effects of humor on memory for non-sensical pictures.* Acta Psychologica, 2009. **132**(1): p. 80-84.
80. Schmidt, S.R., *The humour effect: Differential processing and privileged retrieval.* Memory, 2002. **10**(2): p. 127-138.
81. Coronel, J.C., et al., *Political humor, sharing, and remembering: Insights from neuroimaging.* Journal of Communication, 2021. **71**(1): p. 129-161.
82. Krishnan, H.S. and D. Chakravarti, *A process analysis of the effects of humorous advertising executions on brand claims memory.* Journal of consumer psychology, 2003. **13**(3): p. 230-245.
83. Kramer, C., *I Laugh Because it's Absurd: Humor as Error Detection*, in *It's Funny'Cause It's True: The Lighthearted Philosophers Society's Introduction to Philosophy through Humor*, J.M. Henrigillis and S. Gimbel, Editors. 2021, Lighthearted Open Access.
84. McAndrew, F., *Your Sense of Humor Can Serve as" Social Radar.* Psychology today. April, 2018. **28**: p. 2018.
85. Alfano, M., M. Astola, and P. Urbanowicz, *Having a Sense of Humor as a Virtue.* The Journal of Value Inquiry, 2024. **58**(4): p. 659-680.
86. Cheng, D. and L. Wang, *Examining the energizing effects of humor: The influence of humor on persistence behavior.* Journal of Business and Psychology, 2015. **30**: p. 759-772.
87. Strick, M., et al., *Finding comfort in a joke: consolatory effects of humor through cognitive distraction.* Emotion, 2009. **9**(4): p. 574.

88. Benatar, D., *Taking humour (ethics) seriously, but not too seriously.* Journal of Practical Ethics, 2014. **2**(1).
89. Mark Alfano, M.A., *A sense of humour – even a dark one – is a moral virtue*, in *Available at: https://psyche.co/ideas/a-sense-of-humour-even-a-dark-one-is-a-moral-virtue*. November 29, 2022.
90. Ivanova, A.M., S.N. Enikolopov, and O.V. Mitina, *Sense of humor disorders in patients with schizophrenia and affective disorders.* Psychology in Russia: state of the art, 2014. **7**(1): p. 146-157.
91. Sullivan, K., E. Winner, and H. Tager-Flusberg, *Can adolescents with Williams syndrome tell the difference between lies and jokes?* Developmental neuropsychology, 2003. **23**(1-2): p. 85-103.
92. Granadillo, E.D. and M.F. Mendez, *Pathological joking or Witzelsucht revisited.* The Journal of neuropsychiatry and clinical neurosciences, 2016. **28**(3): p. 162-167.
93. Lyons, V. and M. Fitzgerald, *Humor in autism and Asperger syndrome.* Journal of autism and developmental disorders, 2004. **34**: p. 521-531.
94. Willinger, U., et al., *Cognitive and emotional demands of black humour processing: the role of intelligence, aggressiveness and mood.* Cognitive processing, 2017. **18**: p. 159-167.
95. Greengross, G., *Comedians' Smarts, Humor, and Creativity. How intelligent are stand-up comedians?* December 2, 2013: Available at: https://www.psychologytoday.com/us/blog/humor-sapiens/201312/comedians-smarts-humor-and-creativity.
96. Greengross, G., R.A. Martin, and G. Miller, *Personality traits, intelligence, humor styles, and humor production ability of professional stand-up comedians compared to college students.* Psychology of Aesthetics, Creativity, and the Arts, 2012. **6**(1): p. 74.
97. Janus, S.S., *The great comedians: Personality and other factors.* The American Journal of Psychoanalysis, 1975. **35**: p. 169-174.
98. Feingold, A. and R. Mazzella, *Psychometric intelligence and verbal humor ability.* Personality and Individual Differences, 1991. **12**(5): p. 427-435.
99. Feynman, R.P., *" Surely you're joking, Mr. Feynman!": adventures of a curious character*. 2010: WW Norton & Company.
100. Kudrowitz, B.M., *Haha and aha!: Creativity, idea generation, improvisational humor, and product design*. 2010, Massachusetts Institute of Technology.
101. Asimov, I., *Jokester.* originally published in, 1956.
102. Heinlein, R.A., *The moon is a harsh mistress*. 2018: Penguin.
103. Williamson, J., *With folded hands ...* 1947: Radio Yesteryear.
104. Gelly, S., et al., *MemGEN: Memory is All You Need.* arXiv preprint arXiv:1803.11203, 2018.
105. Garfinkel, B., et al., *On the impossibility of supersized machines.* arXiv preprint arXiv:1703.10987, 2017.
106. Veselovsky, V., M.H. Ribeiro, and R. West, *Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks.* arXiv preprint arXiv:2306.07899, 2023.
107. Denning, P.J. and D.E. Denning, *Artificial stupidity.* Communications of the ACM, 2004. **47**(5): p. 112.

108. Kannan, P., *Another Warning Letter from A.I. Researchers and Executives*. June 12, 2023: Available at: https://www.newyorker.com/humor/daily-shouts/another-warning-letter-from-ai-researchers-and-executives.
109. Yampolskiy, R.V., *Predicting future AI failures from historic examples.* Foresight, 2019. **21**(1): p. 138-152.
110. Williams, R. and R. Yampolskiy, *Understanding and Avoiding AI Failures: A Practical Guide.* Philosophies, 2021. **6**(3): p. 53.
111. Scott, P.J. and R.V. Yampolskiy, *Classification schemas for artificial intelligence failures.* Delphi, 2019. **2**: p. 186.
112. Khatchadourian, R., *The doomsday invention.* The New Yorker, 2015. **23**.
113. Yampolskiy, R.V., *AI: Unexplainable, unpredictable, uncontrollable*. 2024: CRC Press.
114. Ziesche, S. and R.V. Yampolskiy, *Considerations on the AI endgame: ethics, risks and computational frameworks*. 2025: CRC Press.
115. Yampolskiy, R.V., *On monitorability of AI.* AI and Ethics, 2024: p. 1-19.
116. Yampolskiy, R.V., *Untestability of AI and Unfalsifiability of AI Safety Claims.* Αεροπορική Επιθεώρηση (Aviation Review), 2024. **132**: p. 40-55.
117. Yampolskiy, R.V., *On the controllability of artificial intelligence: An analysis of limitations.* Journal of Cyber Security and Mobility, 2022: p. 321-404.
118. Hessel, J., et al., *Do androids laugh at electric sheep? humor" understanding" benchmarks from the new yorker caption contest.* arXiv preprint arXiv:2209.06293, 2022.
119. Weller, O. and K. Seppi, *Humor detection: A transformer gets the last laugh.* arXiv preprint arXiv:1909.00252, 2019.
120. Annamoradnejad, I. and G. Zoghi, *ColBERT: Using BERT sentence embedding in parallel neural networks for computational humor.* Expert Systems with Applications, 2024. **249**: p. 123685.
121. Yampolskiy, R., *Against Purposeful Artificial Intelligence Failures.* AGI-Artificial General Intelligence-Robotics-Safety & Alignment, 2024. **1**(1).
122. Uenuma, F., *20 Years Later, the Y2K Bug Seems Like a Joke—Because Those Behind the Scenes Took It Seriously.* Online: https://time. com/5752129/y2k-bug-history/[Zugriff: 13.10. 2022], 2019.
123. Yampolskiy, R.V., *Behavioral modeling: an overview.* American Journal of Applied Sciences, 2008. **5**(5): p. 496-503.
124. Polimeni, J. and J.P. Reiss, *The first joke: Exploring the evolutionary origins of humor.* Evolutionary psychology, 2006. **4**(1): p. 147470490600400129.
125. McGraw, A.P. and C. Warren, *Benign violations: Making immoral behavior funny.* Psychological science, 2010. **21**(8): p. 1141-1149.
126. Otto, B.K., *Fools are everywhere: The court jester around the world*. 2001: University of Chicago Press.
127. MacAskill, W., *Doing good better: Effective altruism and a radical new way to make a difference*. 2015: Guardian Faber Publishing.
128. Nayrolles, M. and A. Hamou-Lhadj. *Towards a classification of bugs to facilitate software maintainability tasks*. in *Proceedings of the 1st international workshop on software qualities and their dependencies*. 2018.
129. Catolino, G., et al., *Not all bugs are the same: Understanding, characterizing, and classifying bug types.* Journal of Systems and Software, 2019. **152**: p. 165-181.

130. Bundt, J., et al. *Evaluating synthetic bugs*. in *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*. 2021.
131. Cotroneo, D., et al., *How do bugs surface? A comprehensive study on the characteristics of software bugs manifestation.* Journal of Systems and Software, 2016. **113**: p. 27-43.
132. Mosolygó, B., et al. *On the rise and fall of simple stupid bugs: a life-cycle analysis of sstubs*. in *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. 2021. IEEE.
133. Ratliff, Z.B., et al. *The relationship between software bug type and number of factors involved in failures*. in *2016 IEEE international symposium on software reliability engineering workshops (ISSREW)*. 2016. IEEE.
134. Yankovitz, B.-e., A. Kasirer, and N. Mashal, *The Relationship between Semantic Joke and Idiom Comprehension in Adolescents with Autism Spectrum Disorder.* Brain Sciences, 2023. **13**(6): p. 935.
135. Ritchie, G., *Logic and reasoning in jokes.* The European Journal of Humour Research, 2014. **2**(1): p. 50-60.
136. Charina, I.N., *Lexical and syntactic ambiguity in humor.* International Journal of Humanity Studies (IJHS), 2017. **1**(1): p. 120-131.
137. Rigby, L., P. Denny, and A. Luxton-Reilly. *A miss is as good as a mile: Off-by-one errors and arrays in an introductory programming course*. in *Proceedings of the twenty-second australasian computing education conference*. 2020.
138. Musuvathi, M., et al. *Finding and Reproducing Heisenbugs in Concurrent Programs*. in *OSDI*. 2008.
139. Shammi, P. and D.T. Stuss, *Humour appreciation: a role of the right frontal lobe.* Brain, 1999. **122**(4): p. 657-666.
140. Slaby, I., et al., *Direct current stimulation of the left temporoparietal junction modulates dynamic humor appreciation.* Neuroreport, 2015. **26**(16): p. 988-993.
141. *Monty Python's Flying Circus: The funniest joke in the world.* Retrieved August, 1969. **27**: p. 2011.
142. Baldi, P. *A computational theory of surprise*. in *Information, coding and mathematics: Proceedings of workshop honoring prof. bob mceliece on his 60th birthday*. 2002. Springer.
143. Yampolskiy, R., *How to Escape From the Simulation.* Seeds of Science, 2023: p. 2768-1254.
144. Yampolskiy, R.V., *Metaverse: a solution to the multi-agent value alignment problem.* Journal of Artificial Intelligence and Consciousness, 2022. **9**(03): p. 297-307.
145. Bostrom, N., *Ethical issues in advanced artificial intelligence.* Machine Ethics and Robot Ethics, 2020: p. 69-75.
146. Ceglowski, M., *Superintelligence: The idea that eats smart people.* 2016: https://idlewords.com/talks/superintelligence.htm.
147. Kadari, R., M.A. Sarche, and M.J. Krantz, *Fatal laughter.* Annals of Internal Medicine, 2012. **157**(10): p. 756.
148. Reiss, A.L., et al., *Anomalous hypothalamic responses to humor in cataplexy.* PloS one, 2008. **3**(5): p. e2225.
149. Totah, A.R. and S.R. Benbadis, *Gelastic syncope mistaken for cataplexy.* Sleep medicine, 2002. **3**(1): p. 77-78.

150. Lo, R. and T.J. Cohen, *Laughter-induced syncope: no laughing matter.* The American journal of medicine, 2007. **120**(11): p. e5.
151. Hempelmann, C.F., *The laughter of the 1962 Tanganyika 'laughter epidemic'*. 2007.
152. Yudkowsky, E., *Will superintelligent AI end the world?*, in *Available At: https://www.ted.com/talks/eliezer_yudkowsky_will_superintelligent_ai_end_the_world/transcript*. April 2023.
153. Mijic, R., *Solutions to the Altruist's burden: the Quantum Billionaire Trick*, in *Available at: https://rationalwiki.org/wiki/Roko's_basilisk/Original_post*. July 23, 2010.
154. Ossola, A., *Grimes, Elon Musk, and the Supposedly Trauma-Inducing A.I. Theory That Brought Them Together*. May 8, 2018: Available at: https://futurism.com/grimes-musk-rococo-basilisk.
155. Annie, D., *Pilgrim at Tinker Creek*. 2015: Canterbury Press.
156. Bubb, C.C., *The Jests of Hierocles and Philagrius*. 1920: Rowfant Club.
157. Cimerman, A., *GPT-4's Incredible Ability to Recognize and Explain Humor in Text and Images*. March 16, 2023: Available at: https://mpost.io/gpt-4s-incredible-ability-to-recognize-and-explain-humor-in-text-and-images/.
158. Marcus, G., *Can AI Make You Laugh?*, in *Humans VS Machines*. May 16, 2023: Available at: https://podcasts.apple.com/us/podcast/can-ai-make-you-laugh/id1532110146?i=1000613196164.
159. Yampolskiy, R.V., *Turing test as a defining feature of AI-completeness*, in *Artificial Intelligence, Evolutionary Computing and Metaheuristics*. 2013, Springer Berlin Heidelberg. p. 3-17.
160. Yampolskiy, R.V., *AI-Complete, AI-Hard, or AI-Easy: Classification of Problems in Artificial.* 2011.
161. McGraw, P., *Comedians Will Be the Last People to Lose Their Jobs to Robots: Learn from the World's Funniest People*. May 14, 2021: Available at: https://petermcgraw.org/comedians-will-be-the-last-people-to-lose-their-jobs-to-robots-learn-from-the-worlds-funniest-people.
162. Ferreira, B., *Joke Telling Robots Are the Final Frontier of Artificial Intelligence*, in *Available at: https://www.vice.com/en/article/joke-telling-robots-are-the-final-frontier-of-artificial-intelligence/*. March 15, 2016.
163. Scheel, T., C. Gockel, and T. Scheel, *Definitions, theories, and measurement of humor.* Humor at work in teams, leadership, negotiations, learning and health, 2017: p. 9-29.
164. Thorson, J.A. and F. Powell, *Measurement of sense of humor.* Psychological Reports, 1991. **69**(2): p. 691-702.
165. Ruch, W. and S. Heintz, *Test measurements of humor*, in *Encyclopedia of humor studies*. 2014. p. 759-761.
166. Edwards, K.R. and R.A. Martin, *The conceptualization, measurement, and role of humor as a character strength in positive psychology.* Europe's Journal of Psychology, 2014. **10**(3): p. 505-519.
167. Sjöbergh, J. and K. Araki. *A measure of funniness, applied to finding funny things in WordNet*. in *Proceedings of the Conference of the Pacific Association for Computational Linguistics*. 2009.
168. Taylor, J.M. *Computational detection of humor: A dream or a nightmare? the ontological semantics approach*. in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. 2009. IEEE.

169. Chapman, A.J., *Funniness of jokes, canned laughter and recall performance.* Sociometry, 1973: p. 569-578.
170. Westbury, C., et al., *Telling the world's least funny jokes: On the quantification of humor as entropy.* Journal of Memory and Language, 2016. **86**: p. 141-156.
171. Cai, Q., et al., *Modulation of humor ratings of bad jokes by other people's laughter.* Current biology, 2019. **29**(14): p. R677-R678.
172. Martin, R.A., *The situational humor response questionnaire (SHRQ) and coping humor scale (CHS): A decade of research findings.* 1996.
173. Lowis, M.J. and J.M. Nieuwoudt, *The use of a cartoon rating scale as a measure for the humor construct.* The Journal of psychology, 1995. **129**(2): p. 133-144.
174. Poria, S., et al., *A deeper look into sarcastic tweets using deep convolutional neural networks.* arXiv preprint arXiv:1610.08815, 2016.
175. Gao, X., S. Nayak, and M. Coler. *Improving sarcasm detection from speech and text through attention-based fusion exploiting the interplay of emotions and sentiments*. in *Proceedings of Meetings on Acoustics*. 2024. Acoustical Society of America.
176. Howe, W. and R. Yampolskiy. *Impossibility of Unambiguous Communication as a Source of Failure in AI Systems*. in *AISafety@ IJCAI*. 2021.
177. Adebayo, G.O. and R.V. Yampolskiy, *Estimating intelligence quotient using stylometry and machine learning techniques: A review.* Big Data Mining and Analytics, 2022. **5**(3): p. 163-191.
178. Brcic, M. and R.V. Yampolskiy, *Impossibility Results in AI: a survey.* ACM Computing Surveys, 2023. **56**(1): p. 1-24.
179. Yampolskiy, R.V., *Unpredictability of AI: On the impossibility of accurately predicting all actions of a smarter agent.* Journal of Artificial Intelligence and Consciousness, 2020. **7**(01): p. 109-118.
180. Aikin, S.F., *Poe's Law, group polarization, and argumentative failure in religious and political discourse.* Social Semiotics, 2013. **23**(3): p. 301-317.
181. Seewoester, S., *Linguistic ambiguity in language-based jokes*. 2009, DePaul University.