How to Justify the Harms of Offensive Humor

OSKAR TELLHED

Abstract

In this essay I examine the ethics of offensive humor through the lens of the Benign Violation Theory, which posits that humor arises when something seems wrong (a violation) but also okay (benign). I argue that while offensive humor can cause harm by promoting stereotypes and disengaging compassion, ultimately the arguments for free speech justify allowing it, as restricting such speech could lead to a slippery slope of unjustified censorship. Instead of restriction, the better response is more speech and efforts to address the root causes inspiring offensive humor.

Keywords: Ethics, Humor, Benign Violation Theory, Stereotypes, Free speech, Democracy, Censorship

1. Introduction

In 2018, a controversy raged in Sweden. It all started when a group of celebrities vowed for the removal of a particular song from the streaming service Spotify. The song in question, released three years earlier, was called “Knulla barn” (“Fuck/rape children”). The artists behind it were comedian Anton Magnusson, aka “Mr Cool”, and comedian and cartoonist Simon Gärdenfors. As a consequence of the media attention, the song was removed from Spotify, the artists had many of their future gigs cancelled, and were accused of ‘incitement and corruption of minors’ (which inspired the titles of their subsequent standup specials) but the police investigation was ultimately dropped.

Many would probably agree that a lot of humor is ethically uninteresting. Most puns seem rather harmless, as is the case with a lot of observational humor, such as jokes about airline food. There are categories of humor, though, that might warrant our philosophical attention. Racial and gender humor, for example, and jokes about God and religious figures, so called blasphemous humor. Other categories are scatological (genitalia, sex, urination, defecation, menstruation) and

morbid humor (dead baby jokes, making light of the Holocaust, famine and disease) or humor about people's personal attributes, like their big ears or mental or physical disabilities.

Is it morally permissible to joke about these topics? Is there a line one should never cross, and if so, where is it? I am going to argue for the view that the harm that offensive humor can cause is ultimately justified by arguments for free speech.

In order to get there, we will first make sure that we have a clear definition of humor, because until we know *why* we laugh — what humor *is* — we can not make definitive judgements concerning its ethics. Once we've settled on a theory of humor and laughter — the Benign Violation Theory — and concluded that humor's potential to cause harm is very serious and real, we ultimately have to ask: if humor can cause harm, what, if anything, justifies its existence?

In the discussion part of the essay, I will weigh the arguments for free speech against those for the restriction of humor, and in the end apply it to the aforementioned controversial song.

2. The History of Humor

The first to analyze laughter and humor was Plato. And he was not very positive. He linked amusement to moral shortcomings, and since apparently in laughing we lose physical and psychological self-control, and since we express malice towards those at whom we laugh (more on that later), amusement, Plato concluded, produces aggression (Morreal 2020, 632). Christian thinkers continued on this critical path, and the monastery was particularly harsh about laughter. Indeed, in the Bible, the only way in which God laughs is maliciously. And although the Protestant Reformation changed many things, the rejection of laughter was not one of them. Plato's understanding of laughter as malicious influenced Thomas Hobbes, who officially developed it into what is now called *The Superiority Theory*, and states that we laugh when we feel superior to that or them at whom we laugh, not seldom other people's misfortunes or shortcomings. A century later, however, the theory began to be challenged, with the claims that feelings of superiority was neither necessary or sufficient for laughter, like when we laugh at an odd metaphor or clever wordplay, nor do we always laugh when we observe others' misfortune, such as when we see a poor beggar on the street (ibid., 636).

Thus, a new theory needed to be developed. The one most accepted today is called *The Incongruity Theory*. One of the earliest ones to talk about this was Kant (and Schopenhauer subsequently gave it its name.) Kant argued that laughter is the direct effect of unfulfilled expectation: when the punchline is contrary to what we expect it to be, the anticipation is frustrated, and, having no other way to release the energy, we laugh (Marra 2019, 33). Although this is still one of the most popular theories among philosophers, it is the target of serious objections. Most of all, it is not able to distinguish between humorous and non-humorous incongruity; the theory "often predicts that joyous, tragic, and awe-inspiring experiences will be humorous even when they are not." (Warren & McGraw 2016)

Which theory one views humor through dictates the way one judges a piece of humor. If one is (knowingly or not) a Superiority theorist, they would likely see, for example, the 2005 Jyllandsposten cartoons, as a malicious exercise in disrespect, meant to demean in order to feel superior to Islam. If one is an Incongruity proponent, however, the cartoons might be seen as a critical, liberal means of dissent, rather than cruelty and disrespect. As Nicholas Holm (2011) puts it: "When humour is understood as an exercise in incongruity and satire it becomes a freedom of speech issue. When humour is understood as superiority and mockery it becomes a tolerance issue." (Holm 2011, 19)

Why humans laugh is a contested question and has been so for 2500 years, and until we have firmly settled on a theory, we can not make definitive judgements concerning its ethics. How so? Imagine a doctor prescribing medicine without first understanding the disease; just as one needs to diagnose an ailment before prescribing treatment, we need to understand why something makes us laugh before determining if it's ethically acceptable.

There are over a hundred theories for why we laugh, and I will now present the one among them that I find to be the by far most promising.

3. Benign Violations

In 2010, Peter McGraw and Caleb Warren introduced a new theory which they called *The Benign Violation Theory* (BVT). Warren and McGraw sum up their dismissal of the superiority- and incongruity theory (as well as the third of the Big Three theories, Freud's *Relief Theory*) like so:

"Unintentionally killing a loved one would be incongruous and assert superiority, and release repressed aggressive tension, but is unlikely to be funny." (Warren & McGraw 2010)

According to BVT, three conditions need to be met in order to elicit laughter: a situation must be appraised as a *violation*, as *benign*, and the two appraisals must occur simultaneously.

One of the big advantages of BVT is that is can offer an evolutionary account for laughter. When primates are play fighting, tickle each other, or are in the presence of other physical threats, they very often laugh. However, the situation must at the same time be perceived as safe, nonserious, playful — in other words, benign (Warren & McGraw 2010). As humans evolved, developed a sense of self, culture, language, and a system of logic, the situations that could elicit humor grew from only involving physical threats to a wider range of violations, such as threats to identity (e.g. insults), cultural norms (e.g. awkward greetings), linguistic norms (e.g., puns), logic norms (e.g. absurdities) and moral norms[1] (disrespectful behavior; Warren & McGraw 2014c). BVT posits that anything that's threatening to one's sense of how the world *ought to be*, as long as the situation also seems benign, will be humorous.

Let's look at an example. In a study, participants were told this short (and hopefully fictional) anecdote:

> "Before he passed away, Keith Richard's father told his son to cremate his body. Then he told Keith to do whatever he wished with the remains. Keith decided to *snort* his dead father's ashes." (Warren & McGraw 2010)

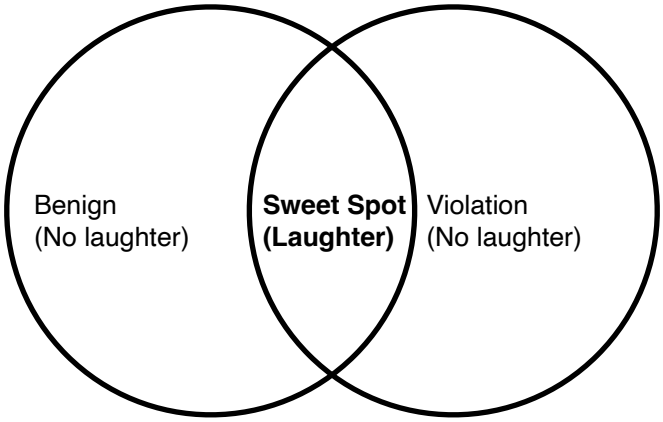
It turned out that participants who interpreted Keith's behavior both as wrong *and* not wrong showed significantly more signs of amusements than those who saw the behavior as strictly wrong *or* not wrong. (ibid.)

So what tends to make something that seems wrong seem also *not* wrong — in other words, a violation seem benign? There are three main reasons: (a) while one norm suggests something is wrong, another norm suggests it is acceptable, (b) one is only weakly committed to the violated

[1] Since we're concerned with ethics, in this essay we will not spend any time on violations of linguistic, cultural, nor logical norms, but solely *moral* ones.

norm, or (c) the violation is psychologically distant. (ibid.) There are commonly four forms of distance: temporal (now vs. then), spatial (here vs. there), social (self vs. other), and hypothetical (real vs. imagined). For example, something disgusting is interpreted as more amusing when it's ostensibly fake, seems far away in space or time, or happens to someone else (McGraw & Williams & Warren 2014b). There's an old saying: "comedy is tragedy plus time." According to BVT, tragedy turns into comedy after just the right amount of time; not too little yet not too much. (ibid.) For example, people report that getting hit by a car would be more humorous if it happened five years ago than if it happened yesterday, while stubbing a toe would be more humorous if it happened yesterday than five years ago (ibid.).

BVT shows how a situation can fail to be humorous in two ways. It can either be purely a violation (say, being tickled by a creepy stranger) or purely benign (tickling yourself) (ibid.). Humor requires you to hit the "sweet spot" in between.



So what does BVT entail for our discussion? Let's go through a few points.

First, it appears to be necessary, per definition, to feel a degree of, at least, *discomfort*, when perceiving humor. Whereas humor has traditionally been seen as something purely pleasurable from a psychological standpoint, this provides that presumption with a twist.

Second, BVT explains why what is seen as humorous changes from generation to generation, culture to culture, person to person. Norms change, and thus what we find to be (or not to be) violations of those norms. As Warren and McGraw (2014c) put it, "This is why a baby farting at

a fancy dinner might seem normal to the baby, hilarious to the rambunctious older brother, and embarrassing to the mother who wants to make a good impression on others." (Warren & McGraw 2014c, 76) Everyone can potentially laugh at anything, as long as it hits their sweet spot. Something can be universally funny, but not *objectively* funny. Humor is, in other words, and perhaps unsurprisingly, highly subjective.

Third, since nothing is objectively humorous, nothing can be objectively *not* humorous. If a racist finds a racist joke funny, it is, whether we like it or not, for all intents and purposes, indeed *funny* (Woodcock, 208).[2]

Fourth, since a humorous event only becomes funny in virtue of its collision with a conscious being's norms, nothing is funny independent of a human's reaction. In other words, nothing is inherently funny in itself. If a joke is told in the woods and no one hears it, it's not funny (except maybe for the teller).

Since it's often impossible to analyze why (within the framework of BVT) a person actually laughs at a joke — whether it depended on a moral violation, logical violation, what served as a distance-maker, whether it concerned one's own norms or those one perceives to be society's, etc. — we should leave the listener be, and henceforth focus on the *speech act of the speaker*. After all, what we're after is an answer to when it's right/wrong to *tell* a joke.

4. The Effects of Humor

Can humor have negative effects on people's well-being? The fact that I'm writing this essay seems to imply that it can. However, not all would agree. Some claim that humor doesn't affect or influence the audience at all, and that it's merely a 'thermometer' of society. One could say that the comedian merely produces words that s/he knows will hit the majority of the audience's sweet spots (i.e. benign violation of their norms) based on calculations including current events, cultural commonalities, etc., and thus merely acts as a thermometer of the audience's society. If everyone were to laugh at immoral jokes, then the problem lies with the audience, and society at large, and not with the comedian, who's not more than a mirror of their values and norms. Thus,

[2] Whether there's even a correlation between laughing at immoral jokes, and being an immoral person, however, is a contested matter, and a subject for another essay.

according to this view, the comedian is utterly harmless; if a thermometer measures high temperatures that kill people, it's not the thermometer's fault, but the heat's.

The counterargument to this, however, is that while humor can be a consequence of norms, norms can also be a consequence of humor. Humor is a neutral thermometer only if one can prove that it never has any negative consequences, which many claim it does. For example, it is said that sexist jokes can contribute to legitimizing sexual misconduct (Taylor & Francis 2017) and jokes that hinge on ethnic or racial tropes often end up reinforcing negative stereotypes. (Yuan 2023) But, one could argue, saying that humor can normalize bad behavior is to underestimate the moral foundation of society, and the capacity for autonomous thinking among its citizens. One could make a comparison to the among religious people common saying, 'if there were no God, everyone would kill'. But as we see in secular societies, that's not really the case. One could also question whether the purported correlation between jokes and behavior has anything to do with *causality.* Do jokes about domestic violence result in more domestic violence? Or does a high level of domestic violence in a society lead to more jokes about domestic violence? Violence against women is a complex issue with many causes, including cultural, psychological and economic factors. To point out humor as a direct cause simplifies the problem and ignores the more significant reasons for the violence. (ChatGPT, 2024)[3]

This may be so, but it doesn't erase the fact that a growing body of psychology research shows that humor in fact can have detrimental effects. For example, upon exposure to sexist (vs nonsexist) jokes, men reported a higher tolerance of gender harassment in the workplace, and at university, recommended greater funding cuts to women's organizations. Most disturbingly, some men even expressed greater willingness to rape a woman after hearing a sexist joke (Ford 2016).

It seems clear, then, that humor can indeed cause harm. So how does it actually do this? I shall classify the two ways in which it can as *directly* and *indirectly*.

[3] ChatGPT was used to generate the final two sentences of the paragraph.

4.1 Direct consequences

Coming back to the speech act, the *direct* way that humor can harm is when a listener is, for instance, triggered by a joke, which might bring up bad memories, like when a victim of sexual abuse hears a joke about rape, or have their self-image hurt, like an overweight person hearing a joke mocking obesity. While these harms are real, I'm going to argue that the main problem when it comes to humor's detrimental consequences lies with the indirect effects, not the direct. Why? Frankly, because a joke-listener always has the opportunity to leave, turn off, stop listening, after having sensed that the humor at offer is not for them. No one forces an audience member to stay in their seat if they don't like the humor, and if they were indeed literally tied to their chair, *that* would be the crime, not the jokes. But what about the first instance of harm? Shouldn't the comedian be punished for telling a joke that an audience member finds hurtful? Let's look at the aims of a comedian. Per definition, a comedian wants their audience to laugh. If no one laughs, the comedian will not have an audience for very long, and thus no purpose to be a comedian. This means that it is in the comedian's best interest to try to hit as many sweet spots as possible, i.e. not offending anyone too much, as that would not result in laughter. If the comedian's goal was to make as many people as possible upset, they would not be a comedian, but merely a provocateur.

So however much a listener is triggered by a joke, at least the (ideal) comedian had good intentions. But, you might say, regardless of their intentions, harm has in fact been inflicted. If you kill a person accidentally, you might still go to jail, and the comedian always knew there was a *risk* that someone in the audience would be offended. Fair. But what about the responsibilities of the listener? You never know when you're going to hear an offensive joke, but when you're willingly going to consume comedy, you should reasonably expect that there's a chance you might be offended. The offensiveness in jokes come as surprises, much like jump scares in scary movies; you might not like it, but you knew what to expect when pressing "play." There's a Swedish[4] saying: "One time is zero times; two times is two too many." To account for the unpredictability in the dynamic between joke-teller and listener, perhaps this principle should be applied to the context of the direct experience of offensive humor.

[4] A similar (and perhaps the source of origin) proverb is the German "Once is never, twice is always."

4.2. Indirect consequences

Let's now turn to the *indirect* consequences, which according to me are more serious. So what are they? For example, if an unattractive person hears a lookist joke, is offended, and leaves, this won't, according to the "one time zero times"-principle, warrant any punishment. But it might have indirect consequences, such as the spreading of lookism via the audience members who stay around, which results in a less welcoming society for the person in question (and many others) despite not hearing any further jokes. Like we already established above, humor can cause harm, and the indirect way in which it does so is, in my opinion, a more pernicious problem, because it often affects innocent non-participants, and is one that's harder to contain, and thus the category with which we'll henceforth concern ourselves.

The question remains, then: if humor can cause harm, what, if anything, justifies its existence? We have now reached the main discussion, where we are going to look at arguments from the side that wants to allow (offensive) humor to flourish, and those who wish for its restriction.

5. Discussion

We have established that humor can cause harm, especially normalizing bad behavior and spreading stereotypes, indirectly harming people who don't even consume the humor. Much of the outrage over the "Knulla barn" song, I take it, was not just because of the direct consequences, i.e. how listeners were potentially triggered by the lyrics, but due to the indirect ones, i.e. that the song might lead to an increase in child molestation. Apart from the title, which is in itself a literal imperative to "rape children", here are some of its lyrics:

> "Do you know what is more rewarding than a chat?
> A really good child fuck
>
> I'm gonna film little boys licking dicks
> And be fisted by grandpa and then win an Oscar

Who will believe you, you are just a little child

I am a grown man, with contacts everywhere" (Mr Cool, 2015)[5]

Based on what we now know about humor, in order to find this funny, one must consider it a benign violation. What is obvious is the violation of norms regarding, to say the least, decency and legality, but how could it be considered benign? Based on my analysis, one benign-making factor is that it is (to some) obviously satire, and thus distanced from the listener due to being hypothetical and not real. Those who find it too offensive (i.e. not funny; a non-benign violation) likely either a) don't get that it's satire, or b) do get it, but still consider it to cross the line as they don't perceive any benign-making factors. Many of them are likely the ones who called for the song's removal from Spotify, and what we're now going to ask ourselves is indeed: what could possibly be the justification for such an offensive creation being publicly available?

Those who believe that the harm of offensive humor justifies its restriction often cite the necessity of respect of the rights and reputations of others (Anderson & Barnes 2023). Making fun of someone's identity (race, gender, appearance, etc.) negates their personhood, and is therefore, they claim, unacceptable (Yuan, 2023). When you entertain stereotypes for fun, your capacity to feel compassion for the marginalized and oppressed is lowered, as amusement is "a practically and epistemically disengaged state of pleasure in which I am not motivated to take action to correct problems." (Morreal 2020, 643)

This sounds quite serious, however, there are those who find that the harm of restriction is greater than the harm of free speech. In a society made up of free and equal citizens, censorship is, according to them, an objectionable form of paternalism and moral authoritarianism, which has no place in a democratic society (Howard 2024). Restriction of speech equals an insult to citizens who should be free to make up their own minds, and should not be tolerated, "even when", as David Strauss writes, "the speech is likely to persuade people to do something that the government considers harmful." (Strauss 1991, 335) For democratic nations, the potential harm of offensive speech is a small price to pay, since free speech not only *promotes* democracy, but, one might say, is *constitutive* of it. In other words, "to be committed to democracy *just is*, in part,

[5] Translated from Swedish with DeepL (deepl.com)

to be committed to free speech." (Howard 2024, 6) Granting restrictions could lead to a slippery slope where the state would have precedence for abusing their power, for example enacting unjust policies such as silencing political opponents. Furthermore, freedom of speech is important because it can help us "tame our illiberal impulses" (ibid., 8); we instinctively want to punish those who hold contrary views:

> "Freedom of speech helps us to practice the general ethos of toleration in a manner that fortifies our liberal convictions. Deeply offensive speech, like pro-Nazi speech, is protected precisely because toleration in these enormously difficult cases promotes a general social ethic of toleration more generally, thereby restraining unjust exercises of state power overall … It is precisely because tolerating evil speech involves extraordinary self-restraint that it works its salutary effects on society generally." (ibid.)

So how does the "Knulla barn" song fare in this discussion? Some would say that it contains none of the "social values" that justify free speech in the first place (ibid.). Also, some, like Ellie Yuan, claim that jokes should be used to criticize wrongdoing and enact change (Yuan, 2023). However, I don't find these arguments to be that strong, since nowhere does it say that humor has any other purpose than to provide pleasure. What if other art forms would have these demands, like music, or scary movies. They don't, so why should humor?

But still, there's no denying that the song caused harm. Can harm ever be justified? J.S. Mill says that it can. But he makes a distinction between "imminently dangerous speech" (Mill 1859) and speech after which there is time for discussion. The former might require restriction, but the latter should not, according to Mill. Hence his example where criticism of corn dealers as "starvers of the poor...ought to be unmolested when simply circulated through the press, but may justly incur punishment when delivered orally to an excited mob assembled before the house of a corn dealer." (Mill 1859, 100)[6] In other words, where there's time for discussion, harmful speech must be allowed. This is compatible with the oft-proposed argument about "counter speech",

[6] A modern example would be January 6, 2021, where there was not much time between Donald Trump's declaring "If you don't fight like hell, you're not going to have a country anymore." (Woodward 2021) and the storming of the Capitol.

according to which one should respond to speech that is harmful not by restricting it, but by arguing back against it (Howard 2024, 17). According to this view, restrictions are never the answer, but rather *better ideas*, and *more speech*, are (ibid.).

6. Conclusion

I have examined the ethics of offensive humor through the lens of the Benign Violation Theory, which posits that humor arises when something seems wrong, but also okay. I then analyzed the potential harms of offensive humor, including direct offense to listeners as well as indirect societal consequences. Applying this to the controversial Swedish song "Knulla barn," I weighed arguments for free speech against those for the restriction of offensive humor.

I find the arguments for free speech to be very strong, in particular regarding the benefits of continuing the discussion, rather than silencing it. One of the defining features of a democracy is that its society is in constant dialogue, ever negotiating its norms and expectations. If we ban some humor today, we would have to ban some other humor tomorrow; the humorous sweet spot is a moving target, and so is the violations, and perhaps is constantly keeping up with the ever changing norms and enforcing restrictions an unreasonable mission that would need far too much resources. According to me, a better use of the state's limited energies would be to focus on measures such as, as Edwin C. Baker puts it: "those directed at changing material conditions in which racism festers, material conditions of both the purveyors and targets of hate." (Baker 2012, 77) I believe that the intended ends of speech bans would be achieved more effectively by aiming at the root causes of inequality, rather than the resultant jests and gags.

References

Anderson, Luvell and Michael Barnes, *"Hate Speech"*, The Stanford Encyclopedia of Philosophy (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.)

Baker, C. Edwin, 2012, *"Hate Speech,"* in Michael Herz, and Peter Molnar (eds.), *The Content and Context of Hate Speech: Rethinking Regulation and Response*, Cambridge, New York: Cambridge University Press, pp. 57–80.

Ford, Thomas E. *Psychology behind the unfunny consequences of jokes that denigrate.* The Conversation (theconversation.com), 2016.

Holm, Nicholas. 2011: *Reassessing the Right to Laughter: Humour, Dissent and the Liberal Imagination*, in AntePodium, Victoria University Wellington

Howard, Jeffrey W., *"Freedom of Speech"*, The Stanford Encyclopedia of Philosophy (Spring 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.)

Marra, Jennifer. 2019: *Humor, Power and Culture: A New Theory on the Experience and Ethics of Humor*

McGraw, A.P. & Warren, C. (2010). *Benign violations: Making immoral behavior funny*. Psychological Science, 21, 1141–1149.

McGraw, A.P. & Warren, C. (2014). *Benign violation theory*. Encyclopedia of Humor Studies,75–77.

McGraw, A.P., Williams, L.T., & Warren, C. (2014). *The rise and fall of humor: Psychological distance modulates humorous responses to tragedy*. Social Psychology and Personality Science,5, 566–572.

Mill, John Stuart, 1859, *On Liberty*, London: John W. Parker and Son.

Morreal, John. 2020: *THE GOOD, THE BAD, AND THE FUNNY: AN ETHICS OF HUMOR*, The Southern Journal of Philosophy, Volume 58, Issue 4

Mr Cool aka Magnusson, Anton & Simon Gärdenfors. Lyrics to *"Knulla Barn"*, 2015

OpenAI. (2023). *ChatGPT* (May 3, 2024 version). https://chat.openai.com/chat

Strauss, David A., 1991, *"Persuasion, Autonomy, and Freedom of Expression",* Columbia Law Review, 91(2): 334–371

Taylor & Francis. *"Do commonplace jokes normalize sexual misconduct experiences?."* ScienceDaily. ScienceDaily, 13 November 2017

Woodward, Calvin (January 13, 2021). *"AP Fact Check: Trump's call to action distorted in debate"*. Associated Press. Archived from the original on January 31, 2021.

Yuan, Ellie. *Offensive humor perpetuates stereotypes, normalizes harmful student behavior* – The Oracle, May 19, 2023.