Center for Humane Technology | *Your Undivided Attention* Podcast
How Zombie Values Infect Society

Tristan Harris: Hey, everyone. This is Tristan.

Aza Raskin: And this is Aza. This is a scene from the classic 1968 horror zombie flick, Night of the Living Dead.

Night of the Li...: ... [inaudible 00:00:14] no evidence of having been partially devoured by their murderers.

Aza Raskin: Why are we playing you a clip from a zombie movie? Well, zombies are a key figure in Haitian folklore as well as pop culture. We sort of all know their story. They bite someone that we love, and it turns them into a creature seeking our brain. We've decided that that's sort of the perfect metaphor to explain something Tristan and I have been thinking a lot about. We call them zombie values.

Tristan Harris: So in this spotlight episode we're going to explain what zombie values are, and then give some examples. Our hope is that by the end of this episode, you'll be able to recognize the zombie values that walk amongst us and help us think through how to upgrade those values to meet the realities of our modern world. So let's dive in. What is a zombie value?

Aza Raskin: A zombie value is an idea or a value that sounds really good and is really good. It points at something that's really worth protecting, but the way we talk about it is no longer adequate. Because the way we talk about protecting the thing the value points at isn't actually protecting anymore, if we just blindly follow it, it will lead to catastrophe. So let's dive in. I think it'll make it much more clearer if we do some examples.

Tristan Harris: Here's one. The answer to bad speech is more speech. We hear this one all the time. Everyone's like, "Oh, well, if that's a bad idea, we want to make sure we answer bad ideas with a debate." We wanted people to have more speech. But now let's imagine the world with synthetic media where I can generate images at scale, fake text at scale, fake arguments at scale. "GPT-4, give me seven arguments for pro-vaccine, anti-vaccine, pro-mask, anti-mask." Flood the zone with these arguments and then just distribute them as maximally as possible to the people we know will get most triggered by them. Is the answer to that world more speech? No. That value is just not adequate to the new world that we're moving into. Tim Wu first called this out in his 2017 article, Is the First Amendment Obsolete?

Aza Raskin: The core concept of that article, I think, is just really clearly articulated. That is, when we first articulated free speech, it was during a time when it was cheap to listen but expensive to speak. That is, there wasn't that much stuff, so it was easy to listen, but distributing your message took a lot. Now we live in a time when listening is expensive, but speaking is cheap. That is to say, it's easy to get

your message distributed around the world to millions of people, but there's so much it's hard to hear. All right, Tristan, if we're in the solution space, what kinds of solutions could you imagine? What is the golden heart of that zombie?

Tristan Harris: Well, it's that we want to make sure that we're, as a society, hearing the synthesis of all the best arguments for and against something very efficiently. In a world, as you said, Aza, if we have a finite amount of attention and suddenly the cost of listening has gone up because everything is a trade off, are we going to listen to those 10 arguments? We're not going to listen to something else? We need to really be efficient. So when I think about speech and media, what I want is a world in which Twitter is ranking for the synthesis of multiple perspectives because that's essentially batching how multiple ideas can be synthesized very efficiently. So a world that is ranking for synthesis is a world that gets at the heart of what the value of the answer to bad speech is more speech, but it's really like the answer to a flood of information is a synthesis of the best arguments efficiently presented, powerfully, concisely, and memorably.

Here's another example. If YouTube and Twitter were ranking by synthesis, let me take an example, so Lex Fridman, he's a famous podcaster out there. He's doing interviews with people. He'll take a voice, a single voice that's like a pro-climate change voice or an anti-climate change voice, a climate denial voice or whatever. He'll interview these people one at a time, and it'll be like a three-hour interview. This is a very inefficient way that's very one-sided. Our brain is going to be asymmetrically persuaded by listening to three hours of either pro or anti-climate sort of views.

The way the attention and engagement economy works is that individual voices are rewarded, the more extreme and one-sided their perspective and more outrage driven it is. Think about a Jordan Peterson or a Bret Weinstein or a Candace Owens, you end up with people who have staked out some position. They will never change their mind publicly from that position because they have learned that they get the biggest audience and the most attention when they take the most extreme perspective. That doesn't reward the kind of behavior, epistemic behavior, the behavior of modeling how do we know what we know and genuinely being in a truth-seeking mindset. That doesn't reward that. So a synthesis-oriented system of ranking information would reward the sort of truth-seeking exchange of ideas.

Aza Raskin: Let's tackle the next zombie value. Fact checking, that's a zombie value. We live in a world of increasingly incorrect information, so the solution is fact checking. We need to check our facts. What's wrong with that, Tristan?

Tristan Harris: Obviously, coming from a world in which people did not do any fact checking, and it was just a world based on yellow journalism and hearsay and rumor and gossip, the world of fact checking is a better world from that prior world. But the

problem again is that that was based on a different technology environment. In a new world with GPT-4 where I can say, "Hey, GPT-4, write me a 20-page research paper on why the vaccine is not safe with real facts and real statistics about people who died, or by the way, write me a 20-page paper on why the vaccine is safe with real facts and real truths and real statistics and pointing out the vested interests of everybody who said one way or the other who was paying them," I can create a very convincing line of facts that is a fact pattern that will persuade you in one direction or the other. All of the facts are true, but it doesn't really fulfill the spirit of what fact checking was really about, which we want a truth-seeking environment. So fact checking, again, has a golden heart or there's a loved one in there that got bitten by the zombie that is no longer adequate to the new modern world.

Aza Raskin: Then if we're going to move into solution space, I think it'd be useful to walk listeners through instead of, "Is it true? Is it false?" is it true, truthful, and representative? So true is whether something is true or false in the world. That's largely the purview of science. That's fact checking. Did this thing actually happen? Truthful is whether the person who is saying the thing believes that it is true or false, so are they coming in good faith? That's truthful.

Tristan Harris: I think the key one there is representative because there are many facts that we can string in a line but that are cherry-picked anecdotes that point in the direction of persuading someone of something that we want them to see. What we really want is how representative those facts are to a broader situation. One other thing, fact checking is like a two-dimensional answer to a three-dimensional world, and true, truthful, and representative is a three-dimensional answer to a three-dimensional world. Imagine if you're getting attacked in three dimensions, but your solution, your defense is only happening in two dimensions, is that world going to work? No. Fact checking is like a two-dimensional defense to a three-dimensional misinformation, truthful, falseful ecosystem. I just made up a word there.

Aza Raskin: Was that truthful or not?

Tristan Harris: I hope so. You could change my mind.

Aza Raskin: I think then if you're starting to think about how would you redesign fact checking for true, truthful, and representative, you can start to imagine it, that instead of just saying this thing was false or this thing is true, you can imagine some kind of graph with what percentage of the population believes X versus Y.

Tristan Harris: How would social media rank information with this three-part framework of true, truthful, and representative? Well, instead of looking at just what's most engaging or what people click on the most or share the most, imagine there's some kind of collective review where people are rating other people in terms of

how good faith those people appear to be. Examples of that are how much they look like they're changing their mind, that they're open to new information, that they're admitting mistakes that they made.

For example, in The Social Dilemma, I said, no one worried about bicycles when bicycles first showed up. What I was really trying to say is that bicycles didn't destroy democracy. But that was a mistake that I made and people called me out on that. Then what will happen is if you're not operating in good faith, you'll like trench up and defend. You'll dig in your heels and you're saying, "No, no, no, this is why I was right." So being in good faith means admitting a mistake when it happens. That's how you know that someone is earnest and sincere. So imagine an ecosystem in which we are ranking information by who are the sincere thinkers who are thinking for themselves, admitting mistakes, and evolving their perspective in public.

Aza Raskin: I'm just imagining a presidential debate where both candidates are doing nonviolent communication. They're like, "Oh, I think I hear you say X. Is that really what you're saying? Oh, this is where I think we have agreement. This is where I think we have disagreement," and where the candidate that wins is the one that includes the other person's viewpoints and makes the strongest case.

Tristan Harris: Well, what I love about that is that then we'd have more trustworthy leaders. Because right now if people aren't communicating in good faith, you can't trust them. How can you trust people that are not communicating in good faith, in a world that rewards good faith communication, rewards leaders who are trustworthy? Which is, by the way, one of the things we're going to need heading into this trust-destroying world of synthetic media.

Aza Raskin: The next zombie value, we should require informed consent. Informed consent is a concept, is a zombie value. Do you want to say why?

Tristan Harris: Sure. This comes from law and legal theory that people should consent to something before they walk into an experience like being filmed or having your data shared or interacting with the doctor. But what if I manipulate the context inside of which I'm getting that consent from you? For example, you load a website. You're in a rush. You've really got to get the address of that thing to go to the place in Google Maps. You try to copy and paste the address, and suddenly there's this popup that says, "Oh, can I ask you for your consent about do you want to allow cookies?" Do you think you're really going to think very hard about that? In fact, I have friends who talk to even the European Union privacy lawyers who've worked so hard to get this informed consent policy with cookies. All of them said that they just kind of skipped and accepted the cookies.

Aza Raskin: There's a huge data brokerage industry in the US where apps that ask for your location just to get you a food delivery end up reselling that data which are used

by marketers to re-target. It's been used as far as adversaries tracking where American military personnel are. It's ridiculous to think that anyone, when they're just using that food delivery app, would have any idea that their location was going to be used that way. So it just names the lie of informed consent. You just don't know.

Tristan Harris: This also relates to the complexity gap, which is that, are people going to read 50 pages, 100 pages of the terms and service agreement? No. There's more complexity in the world than the understanding on the second line that people have about what they're consenting to. As the complexity of the world goes up, people's ability to read through every page of that and consciously think through everything that they're agreeing to is not going to commensurately go up.

Aza Raskin: When I think about what is that golden heart of this zombie value of informed consent is that it's more like it's informed protection.

Tristan Harris: At the end of the day, what we're really looking for here are defaults that have our best interests at heart, like a parent or a teacher. They would need to protect the holistic picture of values that we care about. So instead of putting the burden on the user to read a 50-page service agreement, we want values that automatically protect the best interests. We're seeing examples of this already in the duty of care provisions outlined in both the California Age Appropriate Design Code and in the Kids Online Safety Act here in the US which CHT has endorsed.

Aza Raskin: That's right. It's not about pushing consent onto the end user and say, "It's up to you." It's saying the world is complex. We are going to take the ownership and responsibility for doing that protection. By "we" here, we mean the technology companies.

Democratize everything. This is one of my favorite zombie values because you'll hear it all over Silicon Valley, which is-

Tristan Harris: Especially right now with AI.

Aza Raskin: Democratization is good. Democratize everything. What's wrong with that?

Tristan Harris: Well, the first thing, let's talk about what's right with that. The reason people want to democratized everything is that there's so many people who have not had access to the best technology, the best medical advice, the best creative tools. If AI can democratize access to all of those things, this sounds like an incredibly good thing. We should be democratizing power to more and more hands, especially those who've not had access to that privileged power in the past.

Aza Raskin: Yeah, I'm persuaded. But I think this is actually your line, Tristan, that just because democratize rhymes with democracy, we just smuggle in and believe that it's always good. Tell me about the cases where democratize is bad.

Tristan Harris: Well, if you democratize access to biology weapons or the ability to synthesize explosives, or if you democratize the ability to hack security vulnerabilities and infrastructure and power plants that make people's gas lines safe, that is not a good kind of thing to democratize. What if AI enables some of those things? Democratizing AI means that more people have access to the most dangerous things. The real three-dimensional principle, just to use the 2D-3D metaphor, again, democratizing things is a 2D value. The 3D value is having democratizing power come with a commensurate responsibility, awareness, and wisdom. This is Daniel Schmachtenberger's line that you cannot have the power of gods without the love, prudence, and wisdom of gods. If you have a world where love, prudence, and wisdom are guiding and always matched with the amount of power that we have, that is an adequate way to shape the kind of values of democratizing things. We want rights to go with attendant responsibilities.

Let's talk about another zombie value, which relates to the earlier work on social media, people should use their willpower when using social media. It's up to them to make their own choices. The parents should really be teaching their kids about how to use social media. This sounds like a good value. By the way, we don't not want a world where parents are having a role in their children's lives. We don't not want a world where people take responsibility for using their products. So it's a partial value. Just like zombie values, there's always something there. There's a golden heart, a golden nugget, a loved one that we care about.

Aza Raskin: What I'm hearing you say is it is important that people take responsibility for their choices.

Tristan Harris: But what this might miss is the asymmetry of power, which is to say, if TikTok has a supercomputer trained on what hundreds of millions of human social primates click on and which videos they watch next, it can make more and more accurate predictions about which videos will keep you watching. Is it really about, quote/unquote, your willpower when you keep using that app, or is it really about the supercomputer pointed at your brain, which is to say, the asymmetry of power and information that it has over you?

Aza Raskin: This really gets at that false idea of technology is neutral, or we're just giving people what they want. They wouldn't click it if they didn't like it. They wouldn't watch that video if they didn't like it. They wouldn't return to the app again and again and again if they didn't like it. But that really confuses the distinction between that which people want and that which people can't help but do.

Tristan Harris: Right. So it's the degree of asymmetry between people's own agency and the system in which they're making choices within.

Aza Raskin: There's a whole area of law on undue influence, which is really trying to measure the asymmetry of power in human relationships. I recommend everyone on the podcast go back and listen to our episode with Steve Hassan on cults, where we dive more deeply into understanding what is undue influence and how do you know when you're under an asymmetric power.

Tristan Harris: One other last thing there is on parents, because I often hear this, is that, "Well, it's the parents' responsibility to educate their kids about all this technology." Well, as the technology evolves and moves at an exponentially faster rate, and now your kids are using Discord and Fortnite and Roblox and 20 other new social media apps that you haven't even heard of and we haven't even appraised of in popular culture yet, do you think we want to live in a world where parents should have to bear the burden of knowing the intricacies and feature sets and design of all of those different apps? That's not a tenable situation for parents to take on that burden. It's not that we want parents to not be responsible. I want to make sure I'm clarifying. Parents do have a responsibility. Again, there's an asymmetry in how much responsibility those parents would have to take on to adequately match how fast the evolution of technology is going.

What legislators, I think, need to take away from this is imagine that instead of responding, "This is the parent's responsibility, the parent's responsibility," we actually passed something called the Tech Fiduciary Act, that we turned all technologies above a certain level of undue influence or asymmetric power into having a fiduciary relationship or a duty of care where their job, their principle interest had to be to design in a way that was of the best interest of the development of that child, let's say, for an app that's serving children.

Aza Raskin: I think a core message for parents and educators is to not internalize the thought of, "This is my fault. If only I could teach kids better mindfulness practices or teach kids better digital literacy, we could solve this problem." And we should teach kids better mindfulness practices, and we should teach better media literacy. We should also acknowledge that that is not enough, and we need larger systemic change.

So takeaways. I think one of the core parts of recognizing a zombie value is noticing where conversation gets stuck because... Let's take free speech versus censorship. Both sides have something really important they're trying to protect. They're trying to protect, as we said, the ability to critique the power structure that you're within. You don't want censorship to do that. You also want the ability to have democracy come together, make sense, and make good decisions. We want both of those things to be true. So noticing that there's a kind of... you

get stuck in the 2D world of it's either free speech or it's censorship, that's a tip-off that you're probably encountering a zombie value.

Tristan Harris: I think that the guidance that noticing where conversations get stuck when you keep returning to the same kind of cul-de-sacs or eddies that just don't go anywhere, you'll sit there for the next two hours and the conversation will never evolve, that's usually an indicator that you're reaching the limits of an idea or a value that has lower sort of dimensionality than the situation is actually requiring of us. We also want to honor why these zombie values walk amongst us. It's because usually they're things that we've held dear that have been passed through lineages. The idea that parents need to be responsible for the choices of their children, that's an important value. These are ideas that are important for a reason. But just noticing when they're coming up against a reality that's not working, I think is a thing that we all need to train in and especially the technologists that are making the world the digital habitats that we are living within.

Aza Raskin: Then in terms of, how do you start finding that golden heart of the zombie and revivifying it? It's not an easy process. It takes real synthesis work. As an example, what do those most against masks and most for masks for the COVID-19 pandemic care about and find the same, and what do they both want? They just both want the pandemic to be over. It's that process of finding the surprising places of agreement that is at the heart of finding the golden heart of zombie values.

Tristan Harris: In the world of free speech versus censorship, we all want to get to a world that we can better understand what's true together and that all voices can be heard. The people who care about misinformation is proliferating, who want to censor that misinformation, they just care about not living in a world where people don't know what's really true. The people who are worried about over-censoring information also care about a world where we're censoring really important facts that are part of our truth-seeking process. So if we can identify that we both want to know what's really true in the world, that's a shared value. We can both want the pandemic to be over. That's a shared value. Imagine if our entire digital habitats and our digital infrastructure was all about fulfilling these shared values, we'd stop getting these endless debates.

Aza Raskin: You don't have to be a philosopher to get involved. We always get the question, what can I do? Well, as an educator, you can sit down with your kids and walk through zombie values and have your kids in your classrooms think about what would be adequate solutions. It's super exciting. As a parent, you can do the same thing. If you are a law student, you can do the same thing. That's what gets me really excited about the concept of zombie values. Because it's sort of a funny term. You sort of imagine a value turned into a zombie lumbering forward from the 18th century, the 19th century into the 21st century. If we could wake

those zombies up, find their golden hearts, revivify the things and the values, the ideas that they're protecting, we could return to the land of the living.

Your Undivided Attention is produced by the Center for Humane Technology, a nonprofit working to catalyze a humane future. Our senior producer is Julia Scott. Kirsten McMurray and Sara McCrea are our associate producers. Sasha Fegan is our managing editor. Mia Lobel is our consulting producer. Mixing on this episode by Jeff Sudakin. Original music and sound design by Ryan and Hays Holladay. And a special thanks to the whole Center for Humane Technology Team for making this podcast possible.

Do you have questions for us? You can always drop us a voice note at humanetech.com/askus, and we just might answer them in an upcoming episode. A very special thanks to our generous supporters who make this entire podcast possible. If you would like to join them, you can visit humanetech.com/donate. You can find show notes, transcripts, and much more at humanetech.com. If you made it all the way here, let me give one more thank you to you for giving us your undivided attention.