The Disagreement Problem

Here is a zebra:

Your first thought was probably that this is not a zebra. In fact, it looks more like a white horse, so I'm not sure why this person is telling me it's a zebra. Fair enough, but let's try another one. Here is a horse:

Having fun? Philosophers call this the zebra problem (sort of). The classic version of this problem has something to do with a zoo, but here's the meta version instead: you see the image above (the second one). You think, hmm that sure looks like a zebra. You probably think this because it has black and white stripes and is zebra shaped. So, it's a fair conclusion to assume that you are looking at a zebra. However, I then come along and tell you that it's not a zebra. I say that it's a horse. I tell you that I have simply added on the black stripes using ChatGPT, and

that it's a horse. I say that because you don't know where the stripes came from, the image could be either a zebra or a horse, but you can't disprove it, so you don't know.

The usual purpose of the zebra problem is to get at epistemology. That is, how do you *know* that the image of the zebra is an image of a zebra? While you might have some evidence (stripes, zebra shaped etc), you also do not know that the zebra is not a horse with stripes added on. If that were the case, then we'd be looking at a horse. So, the zebra problem pushes us to consider how we know things.

This essay is not about the zebra problem. Instead, I'm using it here to motivate what I'm calling "The Disagreement Problem." This is my claim: most disagreements happen not because we fundamentally disagree, but rather, because we disagree on what we are disagreeing about, and we don't even know it.

Here is an example using the zebra problem. You say that the second image is a zebra. Again, I say that it is not a zebra, and that it's simply a white horse with black stripes added on. This is infuriating to you, and rightly so, but you don't have any real way to prove that I'm wrong. This is not a very useful disagreement, because if you define a zebra as zebra shaped and having black stripes, and let's say that I define a zebra only with biological proof like DNA, then we are never going to get on the same page simply by looking at an image. So what we are really disagreeing on here is not if the image shows a zebra, but how we *know* what a zebra is.

Now, this is frustrating to you because it seems as though there's no way forward. However, at this point you're so pissed off at me for all my absurd zebra-related claims, that you are determined to win the argument.

One way forward might be to change the disagreement. For example, you concede that there's no way to prove that the image isn't simply a horse with stripes added on. However, this concession doesn't really hurt you. Remember, you've decided that if something is zebra shaped and has black and white stripes, then it's enough evidence to call it a zebra. As such, all you have to do is get me to admit that the image satisfies those two criteria.

First, you ask me if the animal in the image has black and white stripes. I agree, because in my argument I used ChatGPT to add black stripes to the image, so there must be black stripes. Second, you ask me if the animal in the image is zebra shaped. Predictably, I give you some smug reply along the lines of "how do you know if something is zebra-shaped? What does zebra-shaped even mean?"

Now you've got me. You point out that earlier I said that the animal in the image could either be a horse or a zebra, and that we don't know because we don't know where the stripes came from.

If the animal is either a horse or a zebra, and the stripes are the differentiating factor between the two, then the shape of the animal is not the differentiating factor. In fact, the shape of the animal remains the same in the image, with or without stripes. Given this, and given that our options are horse and zebra, then horse-shape must equal zebra-shape. Finally, because the options are horse and zebra, and horse-shape = zebra-shape, then the animal is zebra-shaped in either case. Thus, you have satisfied your second zebra-defining requirement. I am forced to agree.

Pretty cool. You got me to agree that the image shows a zebra in my terms, and now you get to dismiss me as a stubborn devil's advocate. This is a great example of what happens when you get specific about your argument. Here, the reason we disagreed was entirely to do with how we defined zebra. It was much harder for me to disagree with you about categorical things like stripes. So, if you want to try to agree, figure out what you're disagreeing on first.

Now, while this example is good at demonstrating my claim, it doesn't help us much in practice. We almost never want to get as specific as we did above. If we did, we'd end up with a bunch of little subclaims, and it would be hard to ever disagree on anything substantive. Because we want to disagree on substantive things, it's not a good enough rule to say that we ought to always be more specific about our disagreements. Instead, the best way to disagree might be to find some point between the current, broad state of your disagreement, and the most narrow state. Here's a visualization to clarify this idea:

Current disagreement <*————————————> Specific disagreement
(broad) *You are here* (narrow)

Given that by definition you begin at the leftmost point, your goal is to move to the right. So, a more productive solution than staying at your current state or moving too specific, would be to work with the person you are disagreeing with to find a new point. In practice this might be simply talking about what you are disagreeing about, until you can agree on that, thus forming a new point. This is better than jumping immediately to something specific, because you have some chance at avoiding it. That said, if you can't agree on what to disagree on, you might be forced to the specific disagreement, which is still better than the current disagreement. In either case, doing this is better than staying in your current disagreement.

One final piece of practical guidance that will make this process easier is to be charitable. Being charitable is a fancy way of saying give your opponent the benefit of the doubt. This is easier said than done.

In the zebra case, I was determined to make your life miserable. Unsurprisingly, there are a lot of people who do this. I think this is some unfortunate dimension of human cognition. That is, we have some overriding desire to be right. If you go out into the world and start randomly selecting

people, more often than not they'll probably have this trait. It's not necessarily their fault, more so a combination of them being human, and an alarming uptick in polarized media and so forth that's making us all more stubborn. Regardless of why that is true, it is certainly important.

Given that you will inevitably interact with these types of people, it will help to be charitable. Again, these kinds of people don't like being wrong. The nice thing about being charitable is that you don't intend to make them feel wrong. Even if you vehemently disagree with someone's view internally, you can simply listen. You listen, and you let that person express their view, and try to interpret it in the strongest way you can.

Curiously, when you let someone speak, your odds of changing their mind skyrocket. Think of a typical disagreement. Typically, it is abundantly clear right out the gate that you disagree. And if that is the case, then you've immediately put the other person on the defensive. Because people don't like being wrong, they will defend. And so, you've already lost any hope of changing their mind. But if instead you are charitable, and you do not make it clear that you disagree, then they have nothing to defend. In a pragmatic sense, they are letting their guard down, which is precisely when you ought to get clear on what you're disagreeing about.

So that's it, two principles for better discourse. Get clear on the disagreement, and be charitable. If you do those you'll get a lot more out of arguments with stubborn people like (zebra drawing) me.