I just took the moral sense test that my dad has been wanting me to take for so long. I was expecting to feel a huge sense of empathy and almost inability to finish taking it like Emily Drown had, but I realized that actually wasn’t the case at all. I was able to take it very quickly. It was because I was thinking about the huge importance of why autonomous cars will need to run a certain way.

It reminds me of a conversation I had with Yeng not too far back, I’ll post it here:

Yeng:

If you were a train track coordinator, and you had to make a choice:

Steer left and kill one person, or steer right and kill 5 people. Most people would kill 1 person over 5. That theory goes under CONSEQUENTIALISM, which follows the foundation of utilitarianism (you can read more about Mill’s work on that).

How about this scenario, but it’s 2 people versus 10 people. In this case though, it’s 2 people who are under 16 years old, versus 10 people who are convicted fellons. Would you still follow the principles of consequentialism? Maybe you opt to go for something more moral, in the case of the kids, so then you opt for deontological ethics. normative ethical position that judges the morality of an action based on rules.

But instead of being a train track coordinator, you are by a train station, and you see 5 people stuck on the train. But infront of you is a really fat guy, that if you pushed onto the train track, would kill him and slow the train but would spare the 5 people. In this case, you would also be choosing 5 lives over 1. But what would the ethical differences be in this case? Why is it different? And in under what ethical framework would you argue your case for?

Imagine the following scenarios but with a billion times more context and factors; the AI algorithm that we design has to determine and calculate all of these decisions. If humans themselves cant determine what is right or wrong, how can we be absolute in what a machine would decide is good or bad? How do we even design such decisions? Would there be a standard for something like this?

let’s go back to the 2nd dilemma

the fat man and the train tracks

We can translate this to be more specific on self driving car, because it follows the same premise: do you force an algorithm make a direct decision in the interest of consequentialism or what is morally good for the driver? In essence, let’s say you are approaching a red light. There is a pair of children crossing the intersection safely when you spot that there is a speeding driver on course for a rear-end collision. The car you are driving observes that it can make a legal and safe right turn maneuver to avoid the collision. However, allowing the car to go through will mean the children crossing the street will be hit and likely killed. Should the self-driving car be able to make the right turn because it is a legal and safe maneuver? Or should it be held responsible for saving the lives of the children and take the hit?

We can talk about deontological as well, or more specifically the rules as it follows. The argument weighs both side, but whilst consequentialism yields more favorable results on average, deontological would ALWAYS be consistent.

It’s hard to make arguments for and against any scenario, because we as humans have yet to even decide what decisions to make. And unlike humans, where situations are case by case, algorithms are inflexible and so one scenario dictates most outcomes. That is why I believe self-driving cars will not be done in the near future

Not because the technology isn’t there, but because humans have done little to prepare themselves for it.

In 5 years from now I could be absolutely wrong, and hopefully so, but ethical discussions are sorely lacking because there are so many topics we have to first figure out. On machine learning, there is the ethics of determining outcomes (which is also another case by itself because we need to figure out how to properly/ethically source our data). There is also the case of self-driving cars being regulated? or is there oversight on this technology (AKA private vs public stake)

Me:

see in this case, I consider self driving cars to be something that will only phase into reality if humanity is ready and willing to accept them. This won’t happen if we suddenly spring upon them a new idea of this ‘moral machine’ that can make the decisions for who to save and who to kill. More likely, it will only be accepted if it follows a previous model that it can be emulated from. In the case of transportation, it only makes sense to think of a train. A train is a ‘self-driving’ machine that also ‘chooses’ to kill people who step on its tracks. Similarly, with self driving cars there are two things: first, the everyone would think of a car like a train and if the car was just doing its job (the job being driving, not the job of moral decision making) and they were at fault to be in the way then it would be the same argument as a person stepping on train tracks in front of a train. Second, no one is going to buy a self driving car or get into a self driving car if they think that there is ever a time where the car would choose to save an external being or thing over the passenger. Because of this, it only makes sense that the designers would go with the consequentialism approach. The tricky thing here is that consequentialism will get more users for cars and logically make more sense, BUT we as coders now have a choice for the first time to possibly intentionally code deontoligical decisions, which would save more lives OVERALL. Self driving cars are a harder case study because the motivation is going to be more ‘user’ driven. But in other cases (especially in terms of machine learning algorithms) where the ‘bad’ outcome doesn’t involve death but rather social inequality - THATS where deontological ideals can really come into play!

Yeng:

Okay let’s shift to a topic of algorithms then

Targeted advertising, this is a topic you and I are both very familiar about

The consequentialism framework for adverts takes into account the need for revenue for the provider of free or inexpensive servers and content, plus the utility the user might derive from discovering new or proximally available opportunities, and weighs that against the user’s discomfort in having their data shared.

Depending on the weight placed on keeping services available to all, as well as the utility that individuals are perceived to derive from being exposed to the ads that are selected specifically for them, one might argue that advertising is a necessary evil, or even a positive.

The deontological framework against this might focus on user privacy that allow advertisers access to both the data and the screen real estate, and argue that legal collection of that data is ethically permissible, given the user agreements.

There is also a third case of the virtue ethics framework, which could have the user ignoring the clickbait and advertisements as unworthy, or in following links and even occasionally spending money, so as to support the web of commerce.

How do we design an algorithm that can appease all 3 camps? Isn’t the algorithm being used by Facebook currently a mix of all 3? Wouldn’t that be argued as unethical in most cases?

If we had a machine learning algorithm that determines what is best based on what user responds to (AKA more buys or clicks), would that be ethical? And if not, how do we go about ensuring that it is “ethical”? How do we attempt to regulate such a thing that we ourselves have no agreement on the framework in which we argue from?

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

There’s a lot to think about. I think my dad might have been on to something when he suggested crowdsourcing AI ethics at one point. I think that if we want a machine to be as moral as a human, it only makes sense to allow them to get their morals from humans themselves. Statistically speaking, it would provide the best way to appeal to changing morality and ethics of humanity, as well as fix the problem of possible discrimination or baises against minority groups (since the way the crowd sourcing would work could take into account sample size of minority groups and allow them to have just as much control over their own fates as people in majority groups).

There’s just so much to be said on this topic, but I do think that the idea of ethics and morality and computers and machines and algorithms is what I want to go into. This has always been something I wanted to go into, and now I have found my way in. I have a foot in the door and I’m going to do whatever it takes to wedge it open as far as I possibly can to get into this field. I can’t wait to see what the future holds.

It feels amazing to be passionate about something.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Now that I’m done with the tech talk of this journal entry, I think it is a good time to dive into more life stuff that I haven’t talked about much lately.

I did acid with Eric in Escalante, and it was definitely life changing. I feel like I discovered so many new ways to think about life and had so many epiphanies. The biggest epiphanie for me was thinking about everything in life in terms of macro and micro scales. Escalante is very easy to do this in because it is literally just a giant sand box that has been weathered through time. Something that I thought was funny that I said at one point was “How many Escalante’s do you think are in Escalante?”

It made me think though that not just Escalante is an example of micro versus macro scale. So many things in life are the same. When I worry about a small problem like how to get a spider off of the tent, how I react to that can set up my thoughts and my attitude towards how I solve problems at a macro scale.

Am I going to run away from that spider at night and the zip myself up in the tent and call it a night and let the spider best me?

NO!

I’m going to do whatever I can to get the spider as far away as possible, and if it is too much of a nuisance I might have to kill it. Either way, at that point I can then peacefully enjoy my campsite and my tent without having to lock myself away all night or be paranoid.

The same goes for big problems in my own life. If I encounter a rough patch or something that is a little bit out of my control that seems to be controlling the moves I need to take or make, then I’m not just going to let it do so. I’m not going to submit to my eating disorder or my addictions and bad habits. I won’t just keep myself in the same environment day in and day out and think that something is going to change.

NO! I need to make positive actions to get me to where I need to be.

I need to not have any weed in my apartment, unless it’s for a really extra stressful day, and I will label it as so, with possible distractors and relaxers on the packaging that I need to see before opening it all the way to help myself remember how I feel right now, in control.

I need to clear out all of the miscellaneous junk that is in my apartment, I need to battle underneath the bathroom sink and probably throw away most of what is under there. I need to deep clean the bathroom and the shower so that I feel like I live in a clean area, as well as vacuum the stairs and the room and the downstairs.

I need to go through the kitchen cabinets and either donate or throw away all of the tupperwares and pots and pans and cups and bowls and plates that I don’t need anymore.

I need to go through all of my books and trinkets on the bookshelf and determine which ones are actually important for me to have and which ones I can already pack away in a box for mom to fly back with her (along with the tee-shirt box and any other boxes that make sense).

I need to constantly keep the space that I live in clean and the spaces that I eat in clean.

I need to buy the costco meal prep containers and actually prep my meals for the rest of the week every Sunday (or on Mondays if I’m gone for the weekend). I also need to come up with a list of meal prep recipes that I can make and choose from during each week.

I need to keep up with all of my goals, and continue charting out every day and every week so that I can make sure I am accomplishing them.

I need to allow myself rest at least once a week to relax, read, write, hammock, or just to zone out and do nothing.

I need to spend more time loving myself and caring for myself and my body.

I need to practice self love as much as humanly possible.

There were a few other realizations that I had while on Acid. And I will write about them over the course of the next few days, but I think that will do it for today.

It feels great to be organized and to take time to get things in my life in order!