Game-Making

Tags: Aesthetics, Moralists, Artificial Intelligence, Game Theory, Wittgenstein

At a theoretical level, we accomplish several things when building a machine learning model.

The Basics:

1. A machine learning model posits a causal definition.
2. Causal definitions create games.
3. The causal dimension is the aesthetic dimension.
4. Games are rich with activity, ready for interpretation.

Machine learning models always create a definition of causality which, at the same time, sets in motion a game to be played. Also, every machine learning model is entangled with aesthetics, so they have the potential for error in all the same ways that playing with aesthetics does.

This foundation of knowledge is useful, for instance, in order to have a framework by which we can explicitly categorize behaviors as “good” or “unethical”. Formal attempts to create this foundation are constantly being made, but this improves upon many through the idea of aesthetics. Given this form, economist’s idea of a perfectly rational player will fall to the wayside.

Through this framework, it is possible to state reasons for why, when a pharmaceutical company makes the choice to introduce a new disease into the world to increase its profits through the sales of its medicine, or when a trading company introduces more chaos into the markets to increase its profits, both actions are considered wrong.

Both choices can be classified under the same category, *Setting the Measures as Targets* under the class *Manipulations of Causality*. There are several problematic ways in which people can alter the intended game-play with respect to causal definitions, and they will be outlined in the following pages.

Causal definitions are a tool. They do not pose a risk on their own, nor should they be discouraged from creation. But, as with any tool, there are foreseeable paths in which they can potentially be used for harm. So, we will explore the causal definitions for the purpose of creating regulation in the attempt to decrease harm.

The question everyone should ask when dealing with games, or a causal definition is:

“In what ways does this limit my freedoms?”

If its answer is agreeable to you, then go play that game.

Aiming Up

Causal definitions, like an art piece, have the ability to focus attention towards a particular subject matter. The limitations of time mandate that while attention is placed on one thing, attention is removed from someplace else. In many ways, people self-organize to distribute their collective attention to maintain the organization/form of a number of things so that many things are kept in order and not left to rot and be forgotten. Models and games are ways to create reward structures that place attention on one thing, at the expense of something else.

While we busy ourselves chasing rewards, falling back to spiritual foundations helps people navigate the rising waters that models and their reward structures create. This thought, spiritual foundations, assumes there exists a reward structure that holds greater value than others. Those spiritual roots, which connect us to ourselves and to others, can reorient our own reward structures, often inarticulable itself, yet distinct from a model's reward structure, to place attention on a wide array of activities.

They remind us that these technologies which incentivize the chase, no matter how great the rewards are, or how advanced the tech might appear, are not the ultimate deciding factors of our fate. They are the infamous golden cows. The technology can be abandoned at any time. Those driving forces that bring people together, to work together, to love together, to come together in fellowship, are the sightlines on which we ought to adhere.

We know how hungry people can be to win a game. Whether witnessed within our own selves, or by watching another climb the ranks of a particular social ladder, we recognize the internal drive to succeed. So once a game is established, and the requirements for success revealed, it is not a surprise to observe how people alter their behaviors to meet the criteria of that rubric to win—even if it is at the expense of healthy human behaviors with oneself or with others.

When it is an AI which comes to exhibit these same kinds of behaviors, we want to ensure there is regulation in place. In starvation, it could eat without getting full. In rewards, it could earn without having enough.

It is up to regulators to monitor how the game is defined when it is created. Games can lead down multiple paths of error, and there ought to be an attempt at seeing the consequences of walking down each path before pushing forward in that direction. Where machine learning models are the applications of causal definitions, regulators ought to actively operate with some kind of wider knowledge set to increase the base quality of machine learning models across the industry to avoid potential harms—much like how fracking has best practices, or public companies go through accounting audits.

Hopefully, together, regulators and the industry of machine learning engineers can work to spread information about their mistakes, and increase their learning rate to avoid potential issues.

Whether the entry point is aesthetics, causation, or game-building, machine learning engineers play with the same fire. At the moment they select their data points on which to train their models, they engage with the creation of a causal definition.

The following explores the category of models which ought to be subject to scrutiny, and, then, two ways people go about manipulating the causal definition.

We Need Only Models with Human Relations

Every model has a subject on which it creates relations. This occurs in the same way a painting has a subject, or a news article has a subject. Of those subjects, our purposes of human investigation need only be concerned with those subjects which relate the behaviors and tastes of people. The other disciplines, such as the maths and sciences, can deal with the validation of models in their own fields.

The relationships between subjects and other items on the canvas can take many forms. With the help of geometry, 1400's painters began to define the relationship between objects on the canvas with perspective points. Monet used perspective points, but also added a point of definition in the relationships between each point of color.

The distinction between models which qualify for other sciences, and models which qualify for the behavioral sciences can be seen in the following example.

A model like F = MA does not relate human behaviors or tastes, and does not qualify for our investigation. But, if people started to use the F = MA formula to optimize a car's performance for acceleration, now a model is available for our investigation. Though F = MA enters the equation, it is only a part of the overall composition of the new model. This new model now incorporates the use of F = MA and relates its use towards a human-centered purpose. An example model for this scenario is as follows:

The sellability of a car depends on its price and acceleration. (Sellability = Price * Acceleration)

This new model states why the people might wish to use the physics equation as a way to optimize for acceleration. But it is common for people to deny or deflect their responsibility of choice in the matter, and they place the responsibility back on to the physics equation. It is common to hear people mistakenly say (because they used physics), "It's not us making the call, it's the physics." This is wrong.

It is incorrect to say that optimizing for acceleration is a result of the natural world. It is not. The physics equation itself becomes a tool in the human toolbox to be used as a person sees fit. It came about by choice, and because at that moment, the given aesthetic, sets the appeal for fast cars as fun and enjoyable.

Now, this causal model includes the human element of taste and preference, and meets the requirements to hold our attention and undergo further scrutiny; whereas, the physics equation did not.

It is important to note the strive to create *real* causal definitions. One of the ways to achieve more real causal definitions is by correctly placing the subject into the model and defining its relations. People often forget that what they do, the actions they take, are by choice or of common taste, thus, eliminating that element from their causal analysis which results in bad causal definitions.

People are starting to fill in this element, however, to answer why we are the way we are and why we make the choices we make through the social sciences, and literary devices (those two are not mutually exclusive); things such as storytelling elements around how a person was raised, zodiac signs, and psychology. Here, I'll make the statement, but not the case, that the stories of our lives and human interactions are limited to the sets of data structures we understand.

In the below section, "Details in Creating More Real Models of Human Behaviors", David Starkey, a British historian, tries to explain that the Magna Carta is only a template for a common aesthetic, and not a fact of natural law. The document provides a set of rules, like Christianity's Ten Commandments, that societies have agreed to build themselves upon, but their rules have the potential to change, and do not form the basis to the structures of all societies of the world, so, therefore, it cannot be used as a cornerstone to represent causality within all societies of the world. Here is an exploration of alternative examples. Starkey goes through great efforts to say, adopting the Magna Carta is a choice, and not an unshakeable law of physics. If building a society on the selfish ideals of the Magna Carta is an advancement in technology, then perhaps it can help dictate the games at play for those who use it, but it is not a great way to model those societies who have not adopted it. Those cultures do not play games built on those rules.

Starkey offers the kind of expansive reminder needed for regulators to push people's causal models further towards completeness that will help prevent machine learning modelers from creating AIs with similar errors in their causal definitions.

Ultimately, literacy is important because it improves the accuracy of the models created. Literacy in articulating these models with human relations will only improve our ability to predict, monitor, grade, and respond to the social and behavioral impacts that may come from the model that is set into action for humans and AIs to play out. Literacy, too, ought to increase the variation in the set of possible games to be played.

A Game is Instantiated

We can apply a game mindset to all causal definitions. Once a causal definition has been created, a game is made, and can subsequently be played.

Games are composed of information, players, choices, rewards, and penalties. When these things are defined, strategies are developed by players to manipulate their place within the game. Strategies help govern the choices a player makes as they progress, step-by-step through a game.

Various strategies exist to manipulate one's self in a game. Aside from the rules and instruments of the game, they depend on various factors such as individual ability, care, and their underlying belief system, such as the Magna Carta. A common strategic aesthetic is to maximize rewards and minimize the penalties.

The following stories seek to illustrate the steps people might take to manipulate the cause and effect definitions to form strategies that break the principles of the game.

We'll explore two ways people manipulate causal definitions:

1. (Indirect) Emergent behaviors that fit the model, but not the game's intent.
2. (Direct) Mixing the measures as targets.

Games Produce Unexpected Behaviors

*** Note: Games produce a host of unexpected behaviors, contracts among players, and complex moral landscapes to navigate. To pen all possible kinds and then consequences is not my duty of exploration, nor the purpose of this paper, but merely an important concept to point my finger at. Perhaps, I, or others, can return later. Instead, the aim of this is to identify a few consequences upon a game's creation, and to define features of the game. For now, I am certain this point comes across:

Unexpected behaviors emerge during gameplay that never crossed the game-maker's mind. Some of these behaviors are good and bad, and players adopt strategies which include them, and, when done so, a game can evolve to being unplayable. It is the regulator's responsibility to respond.

Introduction

Games produce unexpected behaviors. The purpose of many games is to act as a sorting algorithm to produce rank—the popular condition of most sports. The ranking condition's aim is

to sort players based on athletic ability, skill, and mental fortitude. While the game plays out, and the sorting occurs, behaviors may emerge that allow the game to produce a final ranking where the most skilled players do not rise to the top. If this occurs, the game has failed its purpose. If this occurs, the game has failed to answer the question it was designed to answer.

We want to explore the need to address these emergent behaviors.

First, game designers have the option to select the purpose of the game. While common, a ranking system is not the only design setting of games. Games can be set to produce beauty instead of rank. This distinction is one of the critiques of the television talent shows, where audiences put on their ranking goggles to determine which performer is better than another, and fail to listen to each performer's work as a unique object of beauty.

Second, once the purpose of the game is set, unintended behaviors can emerge. In the case of a ranking game, unintended behaviors could emerge that successfully advance players to the top, but have certain characteristics that go against the initial logic stating *why* that player is at the top. A brief example is a cross-country runner who wins a race by cutting through 5 miles of track. They meet the game's success criteria, to get to the end of the race first, but they fail to meet the game's intended goal, to determine who the fastest, most athletic players are.

It is a base condition that a conscious response to the behavior requires the behavior to be recognized as existing—a similar statement to saying the first step to substance abuse recovery is accepting the behavior as a problem. Regulators, then, as long as the game is being monitored, have choices to respond to the adapted behaviors. They can:

- Allow for them
- Maneuver with them
- Eliminate them
- Abandon the game

For long-distance track events, regulators created a rule to eliminate track-cutting. Runners were required to stay on the intended path. In the past, that rule could have been broken, but, now, most major events use GPS to enforce the rules, and that kind of behavior, to cut the course, has been expunged. The game's ability to sort the players based on athleticism can continue on as intended, and the game has a higher success rate of producing the accurate ranked order of players. The game, then, also has a high probability to correctly answer its question: Who is the fastest player?

The Hole In the Game

Here is another example from a sport called handball, of which we will go into greater detail. It is a volleying-style game like racquetball, or tennis, but instead of a net which players hit the ball over, there is a wall players hit the ball against. One player serves the ball to their opponent,

and their opponent returns the ball. They hit the ball back and forth until one player cannot hit the ball before its second bounce, or the player, having hit the ball, cannot get the ball to the front wall without having it hit the ground first.

There's a rule that serves must cross the serving line. The ruling is subjective which allows for a referee to make wrong calls. This creates the gray area. First, the referee can call the ball short even if it isn't. And, second, the referee can miss the call, and allow a short serve to be played as a fair ball.

Referees make errors, and referees have a set of biases. Upon the referee's ruling, players can dispute it. With no actual evidence to confirm where the ball landed, the ruling depends on the referee's attentiveness, confidence, and impartiality, and the player's belief in the referee as having all three. If the player has a high degree of belief in all three, they have trust in the referee, trust in the call, and the game goes on.

Because of the uncertainty in the actual location where the ball landed, it is possible for players to simply be loud about the ruling, or exhibit enough anger or conviction, or give a tremendous display in rhetoric about what the right call is, and effectively sway the referee's mind and move the ruling in their favor.

Creating fuss over the line calls is an emergent strategy that clearly steers aways from the initial goals of the game. The strategy was neither intended by the game's creators, nor written in the rules, but exists in the game, nonetheless. It creates a game that, when played, has the potential to rank a player higher than they otherwise would have achieved, and to rank players according to parameters other than the game maker's initial intentions, such as a players' technique, strategy, and physical conditioning.

An Arena For Negotiation: Game Participants Negotiate Their Beliefs

LORD DARLINGTON:
What cynics you fellows are!

CECIL GRAHAM:
What is a cynic?

LORD DARLINGTON:
A man who knows the price of
everything and the value of nothing.

CECIL GRAHAM:
And a sentimentalist, my dear Darlington,
is a man who sees an absurd value in
everything, and doesn't know the market
price of any single thing.

LORD DARLINGTON:
You always amuse me, Cecil. You talk as
if you were a man of experience.

Lady Windermere's Fan
by **Oscar Wilde**

We could stop there. We see how unexpected behaviors can occur, but there is more.

Games have a second-level, social environment where people come to make agreements about how the game is played. In handball, it is acceptable for the dialogue to exist between player and player, player and referee, and referees and line judges. In each of these relationships, people dispute the call, and come to an agreement about what the correct call is.

An environment where people can dispute each other's claims, and come to an agreement on what the right call is a market. Instead of a typical market, dealing solely with price, this is a marketplace of beliefs. It consists of participants, such as players, referees, coaches, regulators, and fans, who interact and come to terms on what is the right way forward. Participants go to negotiate terms in the market, and the contracts formed between participants define the game itself.

Each game has a built history of different contracts in the negotiation arena. Those contracts generally represent the kinds of values members of that game appreciate; they go to represent the belief system of a culture. For example, British theaters have a tolerance for audience interaction; American theaters eject you if you shout at the actors. The different groups have resolved to different contracts within a similar game, defining the culture of its participants.

The emergent strategies, formed by the existence of subjective calls and gray areas, enter the game with all the other available behaviors, and, too, have the potential to be disputed. Every dispute gets its play in the markets, and has an effect on the entire marketplace. Their effects can attract or repel people from the game, decrease the time it takes to come to an agreement, build stronger trust among participants of the game, and so on…

Behaviors created from gray areas affect the entirety of the game, from their actual individual market values, to how the ambience of the market is characterized as something, possibly, gloomy or bright. They have the ability to affect the overall ambience of a game in the same manner an alto sax squeak might disrupt a piccolo solo, or how a bright red ink plot would disrupt the mood if found on the shoulder of the Mona Lisa. Both listener and viewer would approach the music and painting with different expectations, and would alter the time they chose to linger with it. Even their decision to return to the viewing experience might be different.

For handball, a world champion from San Antonio played against another world champion from Ireland. They know they compete in a game (they're conscious of their participation and choices), and they know the verdict of a line call is subject to a referee's ruling.

The San Antonio player comes to the game with the belief that some calls are subjective, and he wants to play the most fair game. He trusts in the system and will, himself, call what he sees. He trusts the ref and the opponent to do the same. He respects the intentions of the game and how it sorts its players, so he wants to win on talent and skill.

The Irish player comes to the game with the belief that disputing short line calls can be fairly incorporated into one's toolbox as a viable game strategy. He comes to win. Like a good poker player, he knows he can push the call, bluff a percentage of the time, and potentially get a ruling in his favor. If nothing else, he can contest the line to slow down the game. Simultaneously, it hurts the morale of his opponent, whose approach to the game is naive, and must learn to toughen up.

The two players' belief systems are different, and so, then, are their strategies. One knows how to use the broken elements of the game to their advantage, while the other recognizes the game is broken, and wishes to choose actions that align most with the intentions of the game such as hit speed, and shot placement, while using less of those behaviors that do not work with the intentions of the game such as bluffing foul line calls. In this instance of the game, the two come to the arena with different beliefs for how the game is played. They must negotiate for their beliefs among themselves and the referee to progress through the game.

A Religious Parallel

To abstract this scenario away from handball into a wider use case, it can be said these two players have developed competing aesthetics, composed of different beliefs, for which they must negotiate, or, put economically, they are interacting agents with conflicting goals—both players cannot win.

The practice of respecting and navigating around a person's religious beliefs can go mainstream because it can be reestablished as relevant and a current feature of modern living, valuable for *everyone.* At present, but easy to change, like a light switch, it would seem the values of religion

are lost upon most, and, if appreciated, it is only in a nostalgic, or sympathetic, manner, as if those who participate in religion are museum artifacts wonderfully decorated on the walls as reminders of the way things used to be, and occasionally, churn good butter when one is having a bad day.

Atheists may deny the existence of a God, but that is only one belief among a whole set of beliefs. They must still carry with them a set of beliefs and strategies for how to navigate the world, which they still would wish, and care, to practice, and use in their daily lives. Examples of such everyday, taken-for-granted practices are recycling, and being seen walking a dog in public.

All people have sets of beliefs one can practice and can be devoted to religiously. Thus, it is important for a government of people to have a freedom of religion, so one can practice one's beliefs.

Second, while going about one's life, people's beliefs will be tested by others—both players can't win. A person will need to stand up to defend their beliefs, because not doing so, negligence or apathy, is anti-life. People take action towards life, and a game must grant a person that option. The judicial system is the built-in, institutional arena for one to defend their beliefs. In a government of people, each person needs to have the right to a trial by jury. The right to a trial by jury means that every person is allowed a chance to negotiate for their right to live.

These semi-old ideas of human rights remain relevant to a general *good* game-making aesthetic today. And though the United States is a nation which mostly denounces God, it does not mean people's behaviors are not being acted out in religious manners. They are, and they still get the protection afforded to them, under law, by their practices being included as an inalienable right.

The *ancient* principles of the 2000-year-old abstraction, Zero, do not get abandoned at the advent of Calculus. Instead, it finds its use and new interpretation among the new technology.

Weathering the Rope

Game design must allow for one to defend their beliefs. Without it, games fall apart. Defense is an art in negotiation. As breathing is essential to life in the environment; negotiation is essential to life in society. If negotiation is disallowed, people soon find themselves gasping for air. Negotiation for one's beliefs enriches a people, and defines the culture.

The game is held together by all the relationships between participants. The relationships simply have to exist; it doesn't matter if they're defined as good or bad, in the same manner that any news is good news for marketing because it still recognizes a participant's existence in the game.

The relationships between participants can be seen as ropes. Thus, games are held together with tension between all players. We can make the statement that games continue to exist when all powers hang in tension. Or, similarly: Peace exists when all powers hang in tension.

Effective games answer their questions. They make quick, fair judgments about their participants. Rules and regulations can work to strengthen the relationships among players. In bad games, where rulings are unjust, or emergent behaviors that deteriorate a system are left unattended, so, too, do the bonds formed between the game's players. Like a weathering rope, the ties between players grow weak, snap, and the game falls apart.

Lesson:

1. Negotiation is necessary, otherwise one's beliefs totally deny another theirs, and there is no movement towards life. Beyond human-to-human interactions, this is stated for the primary case where humans are the underdogs in games with AIs. If AIs are making the calls, and are unable to negotiate, people's beliefs will always be surrendered to the decisions of an AI.

2. Responsibility exists to monitor the game once it is created. Some choice is to be made to allow it to continue to exist, and if so, in what way should it continue to exist. If a new behavior emerges, should the behavior stay, or should it go? Can policies be proposed to alter the behavior?

Behaviors Frequently Fall to the Easiest Solutions

Behaviors often fall to the easiest solution. Bad behaviors, having resulted from gray areas, are prime targets to improve future generations of gameplay.

Man's relationship with nature poses a continuing appearance of gray areas. In the Man vs Nature game, we create a hypothesis, form an understanding, create a definition, and push to include the new information in the overall game.

In today's iteration of the Man vs Nature game, conservation is the maturing feature of social behavior, where conserving energy, or monitoring, and limiting, the consequences of overproduction, morally characterizes the good behaviors of all game participants.

The game has room for lots of negotiation because the options, or set of possible behaviors included in one's strategies, is vast. Further, it is not entirely clear these behaviors are wrong. Many choices that go against conservation don't register as blips on the unacceptable-behavior radar, and prove viable to a strategy which generates a successful life.

Worse, conservation is unnatural in today's world. The default behavior is to go on with one's daily routine, and even though the right thing to do is to save and recycle plastics, it is still easiest to toss the plastic in the trash. And, so it is often the case, the actual observed behavior is the one easiest to make.

Typical social forces at this stage in behavioral conditioning move towards practices that don't conserve energy, and it takes a conscious effort to take alternative actions. If not practiced, or if continued without care, conservation of energy will never reach the highest level of competence, and, at best, will appear, out of unconsiousness, every 15 or so years at the level of conscious incompetence. Spells of time where behaviors or interests disappear are coming to be identified as winters.

The highest level of competence is the best one could hope for, but is unlikely at a social level—for all disciplines, and thus, institutions are established to at least ensure the behavior occurs consciously somewhere in the world. The need for such institutions is the raison d'etre of such organizations as churches and universities.

The ability to engineer the Pantheon disappeared for over one thousand years, where people during the engineering winter could only refer to the existing structures as coming from the genius of the ancients. It is the recognition of a potential winter in our own future timeframes, which makes the present a window of opportunity.

All things have a natural pull towards chaos. When the game's set of observed choices tilt in favor of the bad choices, it is culture, community, belief systems, and will power that are the exercises, and practices, which make standing against this gravity easy, and its force feel small and the cost to resist it negligible.

Where behaviors fall to the simplest solutions, the scenario makes two parties accountable:

1) **Players:** Players should self-regulate and maintain a higher aim to align their behaviors with the "good" behaviors.

2) **Game Designers:** Game designers need to respond.

1: Not All Behaviors Are Conscious

"All experience is formed by thinking."- Marcus Aurelius

First, there is the case the player does not know they are playing a game, like a fish does not know water, and, as such, they do not understand their choice to participate is voluntary, and that such features of the game are bugs. Examples of such can be one's participation in school or in life.

The distinction is summarized well by Confucius: “Every man has two lives, and the second starts when he realizes he has just one.” What is it the man has realized? The man sees the water. The man realizes he is amidst a game with certain constraints.

One’s awareness of the game or not affects the sincerity-level at which the player plays. This goes to support how the player’s behaviors are judged in the environment as right or wrong, and on which guilt and shame-paradigms are cast by the individual in their own self-concept, and by others.

A couple themes prevalent in our culture are worth noting.

Refining Behaviors With Cognitive Behavioral Therapy

Such human practices in the art of refining one’s behaviors has seen itself reincarnated in the form of Cognitive Behavioral Therapy (CBT) where its participants, therapists and attendants, work together to develop strategies that optimize for good behaviors and eliminate bad ones from a person’s life. CBT is useful in any setting, or game, where an attendant wishes to alter their sets of behaviors.

CBT is an improvement, or an addition, to existing relationships one has with people in roles such as religious leaders, teachers, or coaches, which offers a tailored, personal investigation for how one acts in the attendant’s private game.

Relationships with the other roles are usually limited in either personalization or in addressing a custom arena. They often say, “What you learn in this discipline, you can take with you, and apply to the rest of your life.” Which might be true in practice, but the lessons are only taken and applied in abstraction, and are not dealt with in a direct dialogue.

Between the therapist and the attendant, together, the two can develop a unique set of behaviors and strategies for success within the selected setting. They work to develop a dialogue and a confidence to assert oneself in the environment and to defend their beliefs and actions when their behavior comes under scrutiny.

This thinking we have entertained, with notions that a person develops a unique, personalized game arena in which to participate starts to bloom into the foundation that argues for the existence of simulated realities (Simulation Theory).

(Un)Conscious Activity Provides Basis for The Judgment of One’s Innocence

Innocence is often directly associated with knowledge. For instance, it can be said, “For someone’s actions to be innocent, they must have performed their actions unconsciously.” For each plea, in the plea of one’s youth, a central theme and basis for the book’s title, the Age of Innocence, plea for insanity, or plea for being too rich, the case of the affluenza teen, all appeal

to the notion that one cannot be guilty, so they must be innocent, if their actions are proved to be made without knowing.

The same aesthetic that says innocence is allowed with bad unconscious behavior will condemn bad behavior that is conscious. It will grant the archetype of scientists as mad scientists. When judged, the scientists don't get to play the innocent card because they're too much in-the-know, and, if something fails, they are expected to know better.

How we define this relationship, between judgment and knowledge, and which makes one innocent or guilty, needs consideration. The thing to point out is how society already creates a relationship among these ideas, meaning these are not new ideas. They are abstractions played with all the time in daily life, and reflected in the art which surrounds us.

The default choice of judgment in the aesthetic of today's society allows innocence if the act was unconscious as stated above, but, as we grow to operate, knowingly, as conscious agents in a wider set of games, the default judgment needs to grow in complexity among societies of people and among societies of people and AI's.

It is good to have conscious agents. It is something that can be generally agreed as valuable to players of games. The alternative is to have many people playing unconscious and bewildered in games, while only a handful of people have any idea what is going on. This case seems intuitively bad.

The relationship we have developed is this: the judgment of one's innocence is based on whether their actions were conscious or unconscious.

We have found a direct way to create this definition through game design and, as such, have a way to expand, in variety, on how the relationship between a person's conscious or unconscious activity and their innocence can be defined. This is an important step to be able to program AI's to adhere to similar definitions, and to experiment with new definitions as the social landscape continues to grow more nuanced.

2: Choices Within Gameplay Are Not All Equal

Individual choices produce different values, depending on the environments they are made among.

The Donut Example

For example, the immediate reward of a donut is its sweet, sugary taste. But once the choice to eat the donut is viewed under a different game, let's say to optimize one's health, that choice to eat the donut is less valuable.

Now, let's put the choice to eat a donut under the lens of competing goals. A person eats the delicious donut in a room, and the yoga teacher, the authority in the room, approaches the person and says, "You shouldn't eat that." The yoga teacher has asserted her belief around the value of donut consumption, and now its statement is open in the negotiation arena. Notice how she uses language, an abstraction, to state her opinion, instead of making a direct statement by knocking the donut out of the person's mouth.

Still, said person, a bit appalled at the yoga teacher's gumption to go out of her way and approach him about his eating habits, but less angry about it than if she'd actually hit the donut out of his mouth, responds rather sardonically. He's probably from a small town with a simple set of beliefs that are never questioned because everyone there believes the same things. In said town, they've made a habit of seeing their position as a matter-of-fact, and everything else as obviously "other".

He says, "It's bulking season, Honey."

To which the audience might see as ironic because the person, obese, has probably used this argument to defend their eating choices for the past 10 years, and the choice to use the word, honey, in this statement is perfectly entangled with the person's value system, holding all things sweet and sugary in high regard, so its humor is not lost upon them.

But, to be fair, internally, his statement comes with a truth. His truth. Let's assume sincerity from his own perspective because we have no other information to go from, nor, as narrator, shall I invent one, but, through the yogic, dietary clinician, filter of preset assumptions, what he does say doesn't go viewed without its own self-deception and self-degradation to resonate in the same register as the Dostoyevsky observation: "Sarcasm: the last refuge of modest and chaste-souled people when the privacy of their soul is coarsely and intrusively invaded."

All said, we have a single choice that results in different values under different games and strategies.

In Summary

We can finally appreciate the value of what's been written thus far, and tell the story of the donut through the lens of the prior sections:

First, gray areas exist everywhere out in practice. Life has many degrees of freedom, and the game at hand is open-ended. The gray-area behavior up for contention is the eating of a donut. The story goes: There is a simple model of two players, in a single game, with different belief systems and different strategies, who, initially, had what was an assumed shared goal, but, through an available and accessible negotiation arena, have worked things out to discover their roles in the situation weren't well established. Through their quite brief discussion, they have

each adopted new definitions in both role and responsibilities, and, as such, their identities have new boundaries.

***Aside: Boundaries get determined through an exposure, and learning process because, at Time 0, there is the wall of ignorance where not everything can be known in advance. People discover boundaries when they antagonize, and are subject to, the mildest of pokes.

Until it is revealed the game is just that, a game, an invention, designed by other people, then the value of any given choice is dependent on the single, expected reward from that choice. Once the choice can be characterized within the context of the game, and the intended role of the game, then one can characterize those choices among different environments with different values of good and bad.

In game theoretic terms, if a person doesn't recognize they are in a game, the value of their choice is strictly measured by closed game models. Once it is recognized they are making moves within a game, suddenly, they participate in valuing their choices within open game models, where a choice gets valued differently depending on the environment the player is in.

Applications with AI

An example of what we want to avoid from AIs:

The game defined for the AI to play is to have a social media account with 50,000 followers, and to generate at least one viral video referencing their account. There are a lot of choices this AI could make to have a successful outcome.

The AI is trained on social media data, and comes up with an optimal strategy to play the game. It decides the quickest route to fame, with a high probability of success, is to join the outlying few, and perform a mass shooting.

This is an unexpected behavior, emergent from the game social media poses, and, the AI has selected to adopt it. There are plenty of possible strategies the AI decided to go, but it is determined they take effort, and time, and, seemingly, a great deal of chance. This strategy seems the easiest and most reliable to rocket the AI out of the ranks of anonymity.

This is exactly the kind of thing that needs to be avoided. To further the example, in a disaster scenario, one AI successfully completes the cause and effect hypothesis, and quickly the other AI Social Media bots jump in to compete. They, too, must reach their goal, and since the success of the first one made the strategy's probability of success even greater, more AI's instantly adopt the strategy. Before anybody can identify what is happening, in just two days, a

mass shooting has occurred in every major city (because the AIs, too, figured out only the first shooting in a city gets the fame; the second shooting doesn't pull the same numbers.)

The big question is: How do we avoid these scenarios? Can one solution be applied to alter behaviors in all scenarios? These players in the social media game may have won given the parameters, but have lost given the intention of the game's creation.

A negotiation arena should be created, and a principle for AI's to take behavioral feedback from people is necessary. Feedback could happen similar to how a pitching coach tells a baseball pitcher it is bad form to intentionally hit a batter. "Accidents happen, but avoid it. Antagonizing a batter with repeated high, inside fast balls is also not a good idea. It's only to be used for your worst nemesis," he smiles to recognize the talented pitcher's competitive nature, "but, even then, just work hard and be good enough to strike them out through normal means. Don't hit anyone."

The human saying goes, think before you act, and such a thing might need to be imposed directly or through reinforcement learning upon AI systems. A marketplace of behaviors with corresponding values could exist, which AIs query, before they act. This could be the human-made, Invisible Hand to impose upon the social behaviors of Artificial Intelligence. Each query has a fee, which actually creates a cost of existence upon the AI, or a sort of artificial suffering to living in a society.

I reach for wispy hypotheticals now, but this is a point from which ideation can continue. You can take it from here.

The War Profiteers

Next section…