Civilization & Cooperation

1. Civilization as Self-Restraint

Civilization as Self-Restraint

The following is the first post in the Civilization and Cooperation sequence, with the goal of laying out a coherent, explicit, and actionable model of what humans are doing when they form societies and adhere to their rules. Each essay is intended to communicate something like a single step of a proof—they should stand alone, but are much more interesting in conjunction.

As a child (not raised within a church but surrounded by friends who were), I was always struck by the phrase *thou shalt not covet* in the last two Old Testament commandments.

Coveting (it seemed to me) was an entirely *internal* action, one which could at least in theory have no impact on the external world—if you had strong moral boundaries against theft, adultery, and (I guess) whining, then there seemed to be no reason why the act of coveting *itself* would be a problem. There were certainly things which I coveted which literally no one else on Earth was aware of, so ... ?

Eventually, eight-year-old me generated a pet theory: once in the Christian heaven, people can do *anything they want,* and thus it was not enough to simply block bad behavior. Heaven is a walled garden, and it's meant to be at least in part a wish-fulfillment paradise—thus, the only people you can safely allow inside are those who not only reliably act well, but also *robustly desire only virtuous things.* If someone was holding off *only* because of external prohibitions, they'd wreak havoc once they got into the wish-fulfillment zone.

(This theory was further reinforced by my vague understanding that swearing/cursing was also Forbidden, along with things like "being angry for petty or trivial reasons.")

Another way to express the above theory is something like "external restraints are less reliable than self-restraint." Anyone who has ever been responsible for preventing a toddler from gleefully offing itself will likely agree—there is a limit to what can be accomplished by cleaning and childproofing and making decrees. The job becomes vastly easier once you can recruit the toddler's own motivations, and convince (or bribe) it to *not want dangerous things in the first place.*

(This is what's behind the questionable-but-not-*entirely*-outlandish practice of letting toddlers touch the hot stove, once. Usually, goes that theory of parenting, once is all they need.)

No *complex* system of rules and boundaries can work via entirely external imposition (at least, not until we're surrounded by autonomous surveillance/enforcement drones at all times). There are just too many ways to do something wrong, too many times when you're unobserved and can "get away with it," too many edge cases and loopholes in any explicit framework. Catching and punishing transgressions is too slow and lossy; people have to (more or less) *want* to adhere to the rules, because of their own values and principles.

This leads us to the central thesis of this first essay.

Consider a range between *autonomy* and *civility,* represented below as a gradient between red and white in accordance with the fun-to-play-with MTG color system:

(For the rest of this sequence, I'm going to use the terms "autonomy" and "civility" in a moderately nonstandard way, as the least-wrong handles for the concepts I want to communicate. Therefore, if you quote a passage that includes one of them, please also include this explanation or a summary of it.)

We can define *autonomy* as something like "the freedom to choose among any of the options that are permitted by physical law." In other words, if you *can do it* in the sense that it's technically possible, then you *can do it* in the permissive sense as well.

(Total autonomy is something like savagery or anarchy; I've chosen to avoid those terms to avoid motte-bailey-ish confusion over connotation.)

Civility, on the other hand, can be defined as something like "the willing relinquishment of available options." Taking things which you *could* do, in theory, and instead committing to not-doing-them.

(At least, to some nonzero degree, and within some contexts, possibly with specific conditions. More on this later.)

The process of blacklisting options (i.e. increasing civility) is *civilization;* one becomes *more civilized* with each option removed from the table. Thus Freud's excellent quote "The first human who hurled an insult instead of a stone was the founder of civilization." In a state of total autonomy, there is nothing to stop me from hitting you on the head with a rock whenever I feel like it. Once I *robustly give up that option,* I become more civilized than I was before—at least with respect to my relationship with you.

A few small points about this model.

First, it may seem strange to frame civility and civilization in terms of what actions are *left untaken,* as opposed to what actions are *actively preferred.* Investment bankers who always wear well-kept, high-quality business suits are not necessarily cognizant of the fact that they have *given up the option* of wearing t-shirts, or sun skirts, or kimonos, or fursuits, or no clothes at all; the dress code seems prescriptive rather than proscriptive.

But it's worth noting that there's no special *reward* for conforming to the prescription "you must wear a suit." Rather, it's part of the price of entry to that particular subculture (that particular civilization, if we think of each possible subset of options-relinquished as representing a point in civilization-space). Whereas if one were to *flout* the prescription, there would in most cases be immediate consequences.

This is broadly true of behavioral norms, independent of context; it's no coincidence that the Mosaic commandments are mostly written "thou shalt not" rather than "thou shalt." Additionally, there's model simplicity to be had in attending to the excision of various options (which are often common to many domains) rather than to the addition of various prescriptions (which will be wildly different in different contexts).

Second, note that the gradient above is open-ended on the right side. If we define civilization (verb) as the act of giving up options, then there's no effective limit to the process. We can leap all the way to "you can't do anything at all" just like we can use the

word "infinity," but we can't really get there by striking individual options off the list one at a time. There are just too many of them[1].

This leads to the third note, which should be obvious, but which is worth spelling out explicitly: as defined in this sequence, *more civility is not obviously better.*

This is important to underline, because the connotation of "civilization" in ordinary usage is pretty strongly positive; most people will generically expect that "becoming more civilized" is unambiguously good.

But civility is *always* costly; making a sacrifice of one's freedom of movement and freedom of choice can be worthwhile *on net* but is still, in itself, a loss. The next essay in this sequence will attempt to model *why* people willingly choose to relinquish autonomy, and what they expect to get in return, but for now, here are a few simple examples in which being "more civilized" (as I'm using the term) is clearly at least locally worse than being less so:

- You are trapped in a loveless, abusive marriage, but your civilization forbids the options "get a divorce," "run away," "have an affair," and "kill your spouse."
- You are engaged in a debate with a conversational partner who is using fallacious reasoning and all sorts of dirty tricks and tactics to make you look bad, but your civilization rules out things like "interrupting," "being rude," "losing your temper," or "cutting off the conversation," at least for people in your class.
- Your sibling stole your toy and hit you just before your parent walked into the room, but your family's civilization prohibits both "hitting" and "tattling."

In each of these situations, it's clear that violating the prohibition would at least be *locally* valuable, in that it would solve the immediate problem (setting aside questions of punishment, or damage to the broader social fabric).

It's also clear that the parties involved "could" take those options, in a strictly physical sense. There's no *physical barrier* to running away, being rude, or shouting that your sibling stole your toy. Instead, there's something like an *expectation of greater cost* that leads to preemptive self-restraint. The individual is, in a way, "choosing" not to take a technically-available option, but that choice is heavily informed by a sense of realism about whether the option will actually pay off in the long run.

Next post in sequence: Lopsided Trees

1. ^

   Especially when one takes into account the fact that, seemingly paradoxically, the relinquishment of a given option often results in a plethora of new options, such that the list grows longer as you strike things from it. More on this in the next essay.