Self-signaling the ability to do what you want

1

In college, I would often find that I had just a little bit too much food. Either I'd cooked too much or the food I'd ordered was just a bit too large, or whatever.

I'm sure many of you are familiar with the feeling of having four bites of food left, wanting roughly one more bite, but knowing that three bites is too few to justify saving the food for later.

(Then you either apply lots of willpower to save the food for later, or you take another bite, realize that there isn't enough food left to save, and proceed to stuff yourself.)

This is pretty much a standard instance of the sunk cost fallacy, where reasoning of the form "I can't just *not eat the food*, because I already paid for it" neglects the fact that the costs are already sunk. In these scenarios, the *only* consideration should be whether or not eating the rest of the food is be better than throwing it away. Your money, which is gone no matter what you choose, shouldn't factor into the decision.

As a student of economics, I understood the sunk cost fallacy well. But extra food didn't *quite* seem sunk: after all, the food would still give me more calories, and even if it made me overfull for an hour or two, it could lead me to have smaller (and thus cheaper) subsequent meals.

Or, at least, that's the argument that my internal monologue would spin up to distract me long enough for my hands to keep shoving food into my mouth.

The counterargument would go something like

> *First of all, many of the calories will be either wasted or harmful if I consume them now. Secondly, the cost of dinner is more dependent upon what's available than how hungry I am. Third, even if the cost of dinner is reduced, it will be reduced by maybe a dollar, and a few hours of discomfort is not worth a dollar.*

But by then, it would already be too late; the food would be gone and I'd be overfull.

2

Failures of this form can generally be fixed by "just not doing that," which in this case entails forcing yourself to stop eating. I don't like that solution, as it requires an application of willpower, and in general, any solution that requires an application of willpower is a stopgap, not a remedy. I much prefer solutions that get all of myself onto the same page,

including the parts that make distracting arguments so they can shovel more food into my mouth while I'm not looking.

(A problem isn't solved until it's solved automatically, without need for attention or willpower.)

The way I solved this problem was by committing to save any amount of leftover food, *no matter how small*. Two bites left? Screw it, get me a take-out box.

Committing to this, and actually doing it once or twice to show myself that I mean business, had an interesting effect.

First of all, it had the obvious effects that I stopped stuffing myself and that I occasionally had three-bite snacks available in the fridge.

But more importantly, credibly committing (to myself) that I would do the right thing *even if it seemed too late* made it much easier to automatically do the right thing.

Roughly speaking, I managed to signal to the part of myself that was worried about food scarcity that it didn't need to distract me in order to squirrel food away, because I would *actually listen to it.* I showed it that I was on its side, via an unflinching willingness to save food (even one or two bites) with a blatant disregard for social norms and weird looks from confused waiters.

And this, in turn, got that part of me onto *my* side. A willingness (and demonstrated ability) to save any amount of food no matter how small eliminated the *impetus* to keep eating when near full. This, in turn, allowed me to actually look at the remaining food and (armed with more experience about which tiny portions of food are actually appreciated later) and decide whether or not to save it.

These days, my bar for how little food I'm willing to take home is quite low, but I'm *also* comfortable throwing food out (if I'm in a rush or if it won't keep well), and I no longer get the feeling that I'm trying to distract myself for long enough to do something that I wouldn't approve of.

3

I occasionally see people hitting the failure mode where they try to apply willpower in order to do a thing (such as only eat half of their sandwich, and save the other half for later) and then fail slightly (such as by taking a bite out of the second half) at which point they proceed to completely ignore the parts of themselves that suggest restraint (such as by eating the entire second half of the sandwich and thereby stuffing themselves).

I refer to this failure mode as "failing with abandon." It seems to me that it's at least somewhat related to a failure of self-signalling: once the initial target is missed, the target itself is completely discredited and ignored in favor of total indulgence.

The technique I'm describing — self-signalling an ability to do the right thing even if it seems too late — can address this failure mode in general.

People might feel strange saving the second half of the sandwich after they've taken two bites out of it, but *if you actually do that a few times* then it becomes much easier to believe that you *can.* The narrative shifts from "well I guess I'm not saving the second half of *this* sandwich" to "I guess I was hungry enough for two more bites, but now I'll save the rest."

As it turns out, you can do the right thing after missing the initial target! Just promise yourself that you'll allow yourself to do the right thing, no matter how late.

4

There's a certain amount of self-trust that comes from making and honoring commitments to do what you want to do *even after* it's "too late" or "no longer worth it." For me, this entails a certain amount of self-loyalty: I'm willing to accept strange looks from waiters in order to save small amounts of food because I'm more loyal to the part of me that is possessive about food than I am to the social norms.

(I expect this is much easier above a certain confidence threshold, such that others say you are "eccentric" rather than "a weirdo." Your mileage may vary. But don't take that as an excuse; I still strongly encourage you to show yourself

that you are able to do the right thing even after it's "too late".)

I have found that there is significant power in signalling to myself that I'm willing and able to do the thing that I want to do, no matter how futile it may seem; that I'm willing to get as close to the target as possible even if I've already missed it. This prevents me from the impulse to "fail with abandon" in the first place.

5

This technique is one facet of a more general mindset that I find quite useful, which is that of "loyalty to the self." I'll touch upon that general mindstate more next week.