

A Quest for the 21st Century

July 23 2024, by Dereck Piché

We are all born with a complicated motivational system embodied as a neural network. Though what makes us act is complicated and varies with time, we must try to communicate it with each other. Many of our pleasures are collectively intertwined, be it directly or by need for order brought only by social cohesion.

Many minds come with a need for fundamental understanding. These try to reduce phenomena into compressed formulations. These compressions are both crucial for personal understanding as well as communication. Now, some things are irreducible to a size which the neocortex can swallow. One of these is the own reward system it is trying to please. Still, this bias for reducibility has caught many great thinkers in desperate statements. The success in finding extremely compact formulations of the fundamental transformations of matter have led some to apply the same mental tools for morality. For some of them, the neocortex even makes the mistake of searching for this non-existent highly compressed meaning by looking both at the cosmic scale and in crude abstraction of others.

The most stubborn among these people come up with funny surrogates. For example: to maximize understanding of nature, to maximize one's role in the social hierarchy, to minimize suffering in the world, to minimize entropy (or equivalently to maximize free energy). These evidently do not translate accurately the richness of even a single person's needs, but the pursuit of these simple ideals lead to wonderful adventures which do satisfy their minds. Still, some listeners take these literally without much thinking about the consequences of optimizing for them directly, which makes them dangerous as stated goals for the human race.

Then if you want to be broader, less myopic, how should you think about ethics? Is there some clear quest which should be achieved by collective effort this century in order to help ourselves and our descendants in reaching good states? I answer in the affirmative.

Experiences that we want to repeat (those which could be classed as good) vary from individual, in time and intensity. We should then aim, as a collective, to maximize human-proximal variance in a sustainable manner.

I have often been fascinated by the medieval era without knowing why, being a scientist in my age. The medieval era permitted extreme experiences, both good (wanting to be repeated by the individuals) and bad (not wanting to be repeated by the individuals), as well as moderate ones. Yet, these extremes never endangered the extended survival of the human race.

This era I am more fearful of because it is highly existentially unstable. It is only in this era which we are growing technology with the possibility of eradicating the entire human race. (Bioweapons and Unaligned AI, for example). These developments call for retroaction. Now is not a time for moderation. It is crucial for intelligent men and women to come up with solutions that will ensure a human centred highly variable and sustainable future.

A naive approach would be to politically regulate strongly. This does not lead to a variate future and is probably not sustainable (offensive will prevail over reactive technology). A possible solution is to simulate the physical world such that no matter the mayhem caused inside to fill individual ambitions, the outside world remains safe. This is more promising, but risks not being sustainable in the human-proximal aspect. We were not made to be the own gods of our realities.

The ideal would be an Artificial Intelligence game master. Chess remains fun because rules bound the plays such that it allows varied strategies whilst keeping an ethos. Physics and primitivity bound our game for thousands of years. But now that technological primitivity is leaving, physics on its own will not be enough to keep the plays interesting. We need an AI supervisor, which does not ruin the variance and extremes of the human story, but prevents it from being irreversibly destroyed. An AI which lets drama, adversity and disruption happen as long as human-proximity and long-term variance remain. This embodied law, once created, would only act in these special circumstances. The rest of the time, it would simply be a spectator. Our quest is the creation of this regulator. It is one of the hardest and most important quest which we shall ever accomplish.