▫ Let $\mathcal{W}$ be the set of all possible configurations of the universe.

It is common to assume that agents have a fixed utility function $U : \mathcal{W} \to \mathbb{R}$. This is very convenient because it allows us to compute things such as expected utilities, and real numbers allow us to embed preferences in such a way that we can encode any arbitrary trade-offs. Moreover, we have theorems [**?**] that state that any agent that obeys certain "rationality" axioms always behaves as though it is maximizing utility — and so it is reasonable to assume

But taken litterally, this utility function is wrong for a number of reasons (some trivial, some less so):

1.

Most importantly, there is interesting substructure to worlds. We can model this by representing the world in a different way. One common thing to do is assume that there's some set of variables $\mathcal{X} = \{X : \Omega_X\}$, where each variable X can take values in Ω_X, and so we have

$$\mathcal{W} = \prod_{X:\mathcal{X}} \Omega_X$$

The assumption that $\mathcal{W}$ *litterally is* the product of these variables usually reflects modeling choices, rather than the state of the world. In a trivial way, to say that the world can be encoded with a number of binary variables is correct, but totally does not square with the way that these variables are used, in CP-nets, for example.

For instance, if my world $\mathcal{W}$ is over images, then maybe there's some variable `width` and another one height, and one for each pixel (the number of which depend already on two of our variables), so the decomposition already doesn't look quite right. So let's restrict to 1024 by 1024 pixel images. Now there are 2^{20} color-valued variables according to each of the pixels. People might have preferences over images (say, which ones they'd like to hang in their bedrooms), but almost certainly not over the pixels values. Moreover, there is no natural acyclic CP structure on the pixel variables, since really the only time you have preferences over pixels is in context, with a good portion of the image already constructed.

Instead, consider that rather than a world which is litterally the direct product of some variable spaces, we think of the variables $\mathcal{X}$ as features $\{X : \mathcal{W} \to \Omega_X\}$. Mathematically, this buys us nothing, but allows us to decouple the state of the world and its representation from the variables we have preferences over.

In accordance with the observation that preferences over most variables are not actually specified,

1 Going from desires t

2 Adversariality