The general idea is to generalize orders by including extra data. Including a magnitude, for instance, gives us a notion of the strength of a comparison; one can also include reasons, properties.

A

Different aggregations

1 Values

People often model an agent's values by specifying a utility function, loss function, rewards, binary preferences, or goals. In each case the goal is somehow morally the same, but the formalisms are quite different.

Definition 1.1. A *value* (on a random variable X) is a map $\nu : X \to \mathfrak{O}$ where $\mathfrak{O}$ is a ($\mathcal{V}$-enriched) category.

Such a map allows us to think of X itself as a $\mathcal{V}$-enriched category, by

$$\begin{aligned} \mathrm{ob}_{\mathbf{X}} &:= \mathcal{S}_X \\ \mathrm{Hom}_{\mathbf{X}}(x, x') &:= \mathrm{Hom}_{\mathfrak{O}}(\nu x, \nu x') \end{aligned}$$

where composition of arrows and associativity can from $\mathfrak{O}$ can be used directly, with no coherence issues. We think of the hom object $\mathrm{Hom}_X(a, b)$ as tracking the ways a and b should be compared.

1.1 Orders

Example 1.1. In particular, if X is already an (enriched) category, then the identity map id_X is value, sending each $x \in X$ to itself, recovering the additional relations between elements. △

Of course, this is not very useful for articulating the order in the first place — but it means that we can write down a value for X by giving it suitable morphism structure.

Example 1.2. Any pre-order $(O, \preccurlyeq)$ can be regarded as a $\mathbb{B}$-enriched category, where

$$\mathrm{Hom}_O(a, b) = (a \preccurlyeq b) = \begin{cases} 1 & a \leq b \\ 0 & \text{otherwise} \end{cases}$$

The monoidal product is conjunction, so composition $\circ_{a,b,c} : \mathrm{Hom}(b, c) \otimes \mathrm{Hom}(a, b) \to \mathrm{Hom}(a, c)$ guarantees that $a \leq c$ whenever $a \leq b$ and $b \leq c$, and nothing otherwise.

In an order, there is nothing else we can use to compare — from the perspective of a, the only information about b provided by the order is whether or not $a \preccurlyeq b$. △

Example 1.3. In particular, a (naïve) utility function $u : X \to (R, \leq)$ is a value, where the real numbers have their usual order. If X is countable, then any total pre-order $(\leq) \subseteq X \times X$ on it can represented in this way. △

Example 1.4. A goal is a value $X \to \mathbb{B}$. Because $\mathbb{B}$ itself is an order, it is a $\mathbb{B}$-enriched category. △

Of course, there is a lot more to goals (and actually what I've just described here I will later want to call a desire instead), but the general mathematical structure is there.

1.2

We can also track other information aside from just the order. The utility example above, for instance, is somewhat unsatisfactory because the actual geometry of the space is not used. After all one of the biggest reasons to use a utility function at all is that it can capture the *magnitude* of difference between alternatives, which is incredibly useful.

Fact 1.1. *Any group G can be seen as a G-enriched category, where the monoidal product is the group product:*

$$\mathrm{Hom}_G(a,b) = b * a^{-1} \qquad\qquad r \otimes s = r * s$$

and composition works out exactly as one would hope:

$$\begin{gathered} \circ_{a,b,c} : \mathrm{Hom}(b,c) \otimes \mathrm{Hom}(a,b) \to \mathrm{Hom}(a,c) \\ (c * b^{-1}) * (b * a^{-1}) \mapsto c * a^{-1} \end{gathered}$$

Example 1.5. In particular, $\mathbb{R}$ is an $\mathbb{R}$-enriched category, where the tensor product is addition. As a result, a utility function $u : X \to \mathbb{R}$ is a value on X, where $\mathrm{Hom}(x, x') = u(x') - u(x)$. [1] Since we only ever consider differences of utilities, we don't actually need $\mathbb{R}$, just the affine space we get by enriching by itself. △

We have now effectively reduced a utility to a total order with magnitude annotations. Can we then go backwards and re-construct the function, up to a base point?

So far we haven't done anything interesting: we've taken standard things and somehow encoded them in a way that's consistent with a strange definition.

1.3 Yoneda Embedding

One pretty glaring difference between the standard presentation of utility functions and preferences is the number of copies of the variable X.

		Enrichment	
		$(\mathbb{B}, \Rightarrow, \wedge)$	$(\mathbb{R}, \leq, +)$
Shape	1	$X \to \mathbb{B}$ (goal)	$X \to \mathbb{R}$ (utility)
	2	$X \times X \to \mathbb{B}$ (preference)	$X \times X \to \mathbb{R}$ (matrix)

So far we've dealt with this by converting everything to shape 2, which is the easy, boundary-like operation—it's easy to take a utility function u and create a matrix by $\mathbf{X}_{i,j} = u(j) - u(i)$, for instance—but there are some flaws with this approach:

1. It doesn't generalize to other shapes very well — for instance, it's not clear how to encode a selection function which takes any subset of X and returns a supremum (or distribution over its arguments).
2. We eventually want to get to probabalistic transitions, and so while we will get to control the representation and points of the variables, we will ultimately want our arrows to be markov kernels, and so we may not get this level of control directy over the morphisms

[1]Moreover, since addition is symmetric and functorial over $\leq$, $(\mathbb{R}, \leq)$ is a symmetric monoidal pre-order, and so products work out in the usual way [**spivek-ex2.74**].

Instead, we will show that we can build this structure "point-wise", by taking considering pre-sheafs over our $\mathcal{V}$-category, under the yoneda embedding $よ : \mathcal{C} \to [\mathcal{C}^{\text{op}}, \mathcal{V}]$. Note that this is just the curried hom-functor. By the yoneda lemma, natural transformations out of the functor category $[\mathcal{C}^{\text{op}}, \mathcal{V}]$ are in natural bijection with the morphisms of $\mathcal{C}$, and so this will also give us the same enrichment structure.

Example 1.6. Suppose we want to give X a pre-order, by only specifying data at the points — then the yoneda embedding is a value on X, given by

$$\begin{aligned} よ_X : X &\to [X^{\text{op}}, \mathbb{B}] \\ x &\mapsto (\ \cdot \ \preccurlyeq x) \end{aligned}$$

$$\begin{tikzcd} x \arrow[d] & \mapsto & y & \mapsto & x \preccurlyeq y \arrow[d] \\ x' & & y' \arrow[u] & & x' \preccurlyeq y' \end{tikzcd}$$

We write the final line in this way because $\mathbb{B}$ is $\mathbb{B}$-enriched, and so the action on arrows can be read as "if $x \preccurlyeq x'$, $y' \preccurlyeq y$, and $x \preccurlyeq y$, then $x' \preccurlyeq y'$".

Equivalently, it can be thought of set-theoretically: $[X^{\text{op}}, \mathbb{B}]$ on objects, is just subsets of X, and the yoneda embedding encodes an element $x \in X$ as the set of all elements smaller than it. △

To make this even clearer, we can do a very simple case of the above:

Example 1.7. Suppose A is the total order with three elements generated by $a_0 \prec a_1 \prec a_2$, and notate a function with co-domain A as a vector, e.g., $[x, y, z]$ being a map $(a_0 \mapsto x, a_1 \mapsto y, a_2 \mapsto z)$. Then the yoneda embedding gives us a value:

$$\begin{aligned} よ_A : A &\to [A^{\text{op}}, \mathbb{B}] \\ a_0 &\mapsto [1,0,0] \quad \cong \{a_0\} \\ a_1 &\mapsto [1,1,0] \quad \cong \{a_0, a_1\} \\ a_2 &\mapsto [1,1,1] \quad \cong \{a_0, a_1, a_2\} \end{aligned}$$

The action on arrows gives us a natural way to compare them as well. If $f, g : A^{\text{op}} \to \mathbb{B}$ are vectors of this form, then a natural transformation from f to g consists of witnesses that $f(a) \preccurlyeq g(a)$ for all $a \in A$. The yoneda lemma tells us that those natural transformations correspond exactly with inequalities in A.

△

We can also embed our real numbers this way, recovering the affine space of points on $\mathbb{R}$. Restricting to the image of a discrete set, we can visualize this better — carrying our simple example over, we can also get back our magnitude information:

Example 1.8. Now consider a utility function on A, defined by $u(a_0) = 1, u(a_1) = 3, u(a_2) = 4$. The Yoneda embedding of A, regarded as an $\mathbb{R}$-enriched category, is really just curried subtraction

$$\begin{aligned} よ_A : A &\to [A^{\text{op}}, \mathbb{R}] \\ b &\mapsto \quad a \mapsto \big(u(b) - u(a)\big) \end{aligned}$$

We can also see a bit more clearly why the op is there: the reversed ordering on the second argument is reflected in the negation. Taking the suggestive notation from the previous example a bit further, we can get a matrix

$$\begin{array}{cc} \begin{matrix} a_0 & a_1 & a_2 \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 3 \\ -1 & 0 & 2 \\ -3 & -2 & 0 \end{bmatrix} & \begin{matrix} a_0 \\ a_1 \\ a_2 \end{matrix} \end{array}$$

where the columns are the vectors associated with each element of a. Natural transformations from one vector to another, parameterized by $\mathbb{R}$, are additions of constant vectors. For instance, by adding 2 to each component of the middle column, we get the last one, which once again by the Yoneda lemma happens exactly when there was a morphism in the original category.

$\triangle$

We can also encode selection functions this way — that is, an arg max function $2^X \setminus \varnothing \to X$, which selects a maximal element of a non-empty subset.

Example 1.9. This is easiest do do individually for each $k \in \mathbb{N}$, which assumes that X is countable. $\phi : X \times X^k \to \mathbb{B}$ $\triangle$

1.4 So what can you do with them?

Whether or not you can represent things is only half of the picture. It is also important that we retain all of the important features of values that we had before.

Definition 1.2. A value $\phi : X \times A \to \mathfrak{O}$ marginalizes out to a value $\nu : X \to \mathfrak{O}$

- **strongly** if

$$\phi() \simeq$$

2 Beliefs