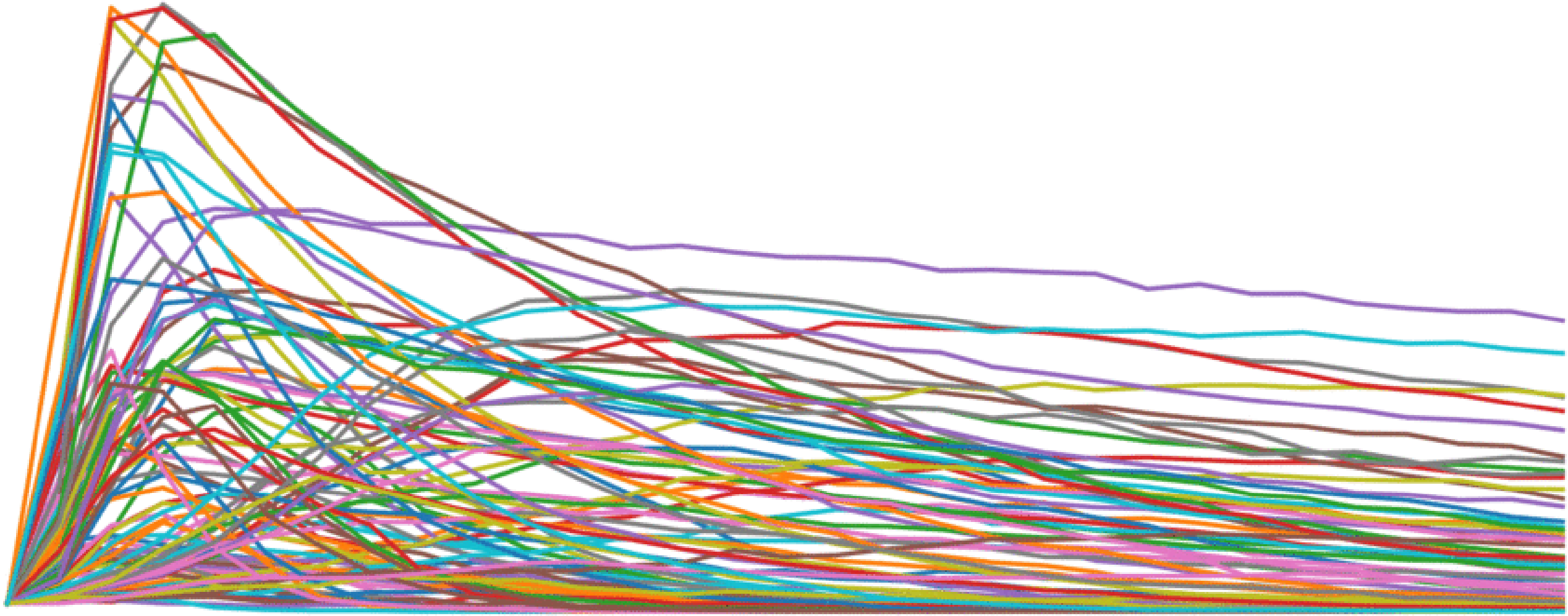
SENSITIVITY ANALYSIS MADE EASY



Who are we?

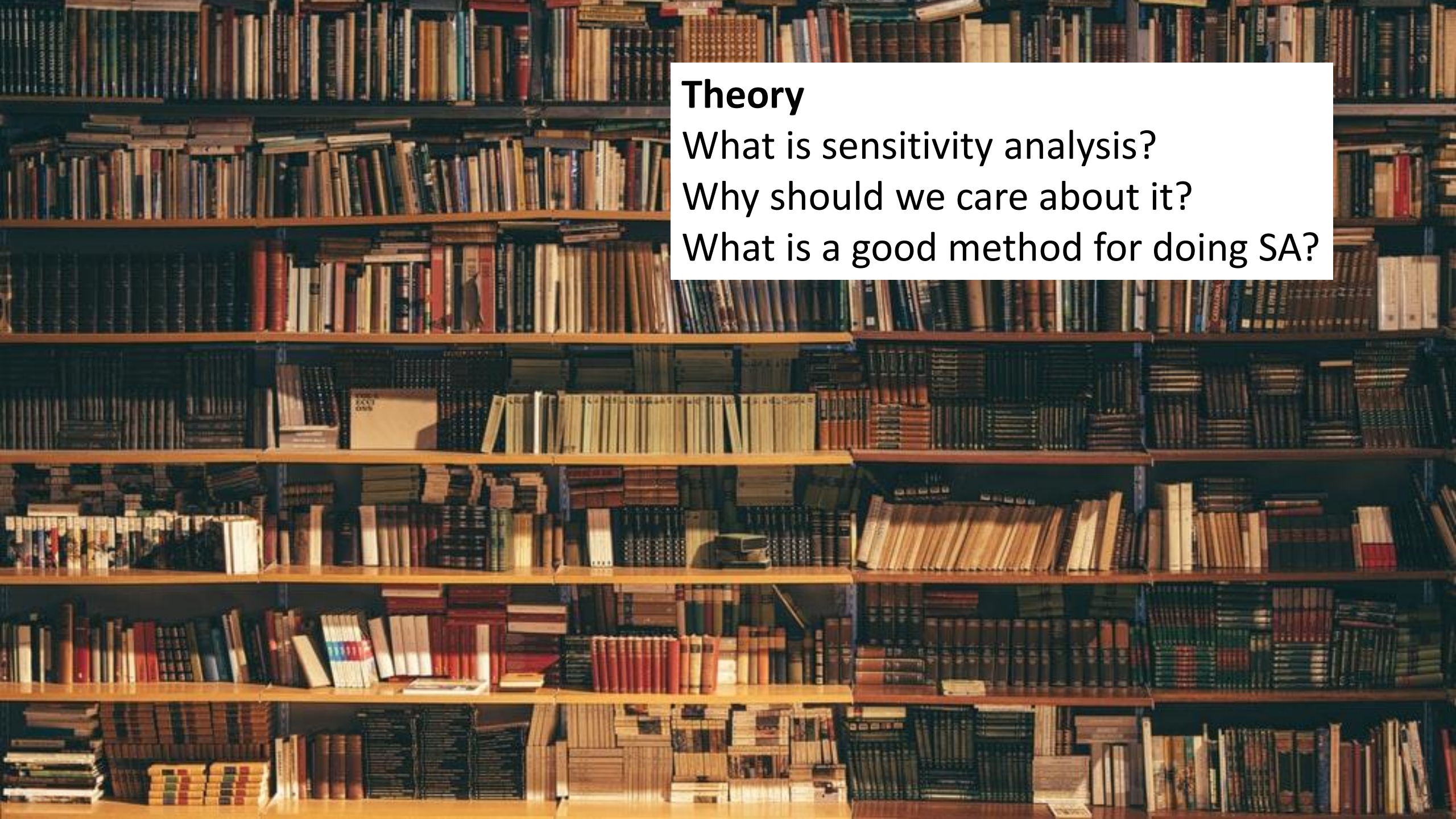
Patrick Steinmann
Wageningen University & Research
patrick.steinmann@wur.nl
@steipatr

Mikhail Sirenko
Delft University of Technology
m.sirenko@tudelft.nl
@mikhailsirenko

Raphaël Klein
Delft University of Technology
r.klein-1@tudelft.nl
@RaphaelKl1

Today's Workshop

1. Theory: what is sensitivity analysis, and why should I care?
2. From Theory to Practice: So I want to do sensitivity analysis, now what?
3. Practice: Let's get our hands dirty!



Theory

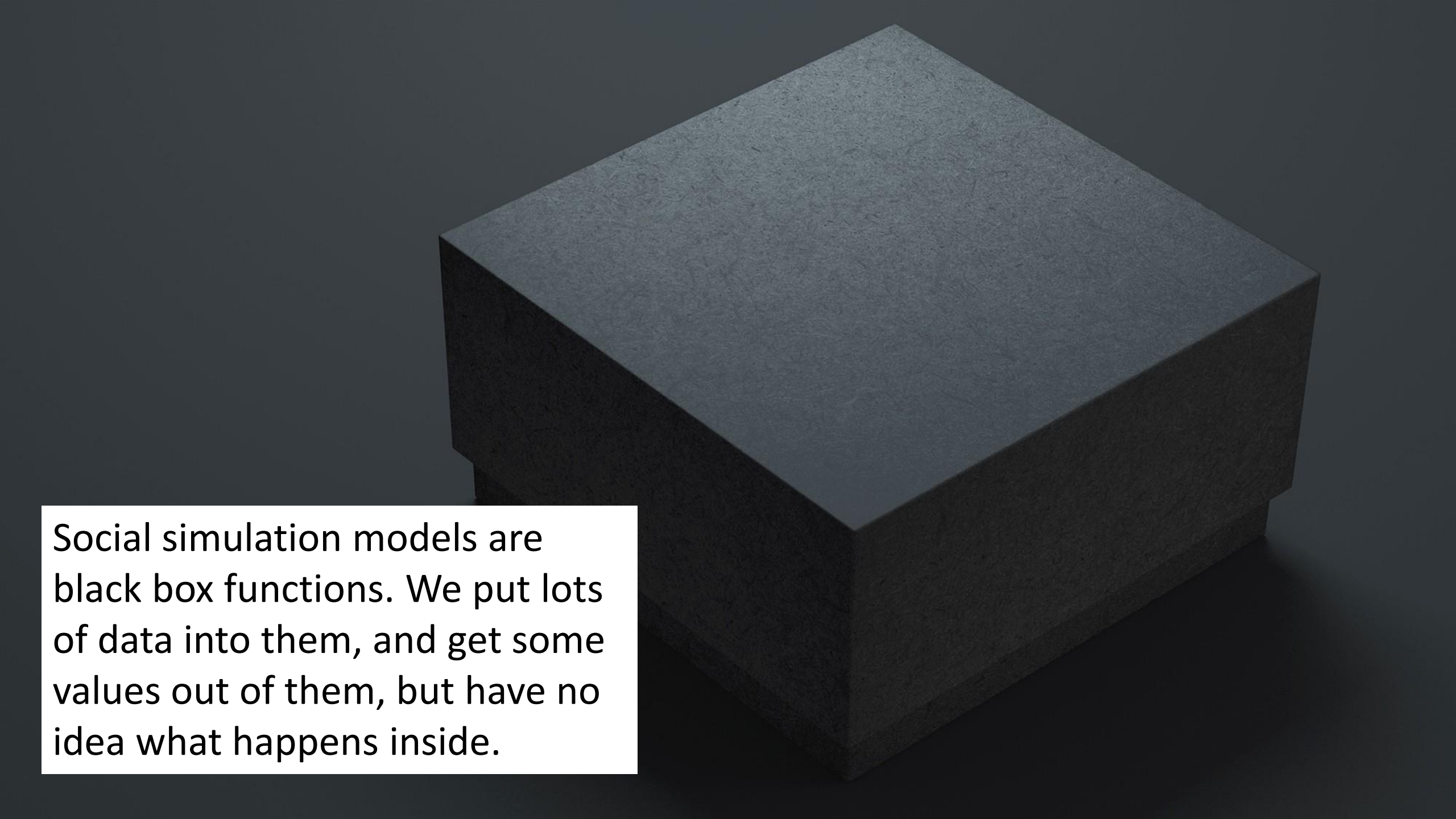
What is sensitivity analysis?

Why should we care about it?

What is a good method for doing SA?

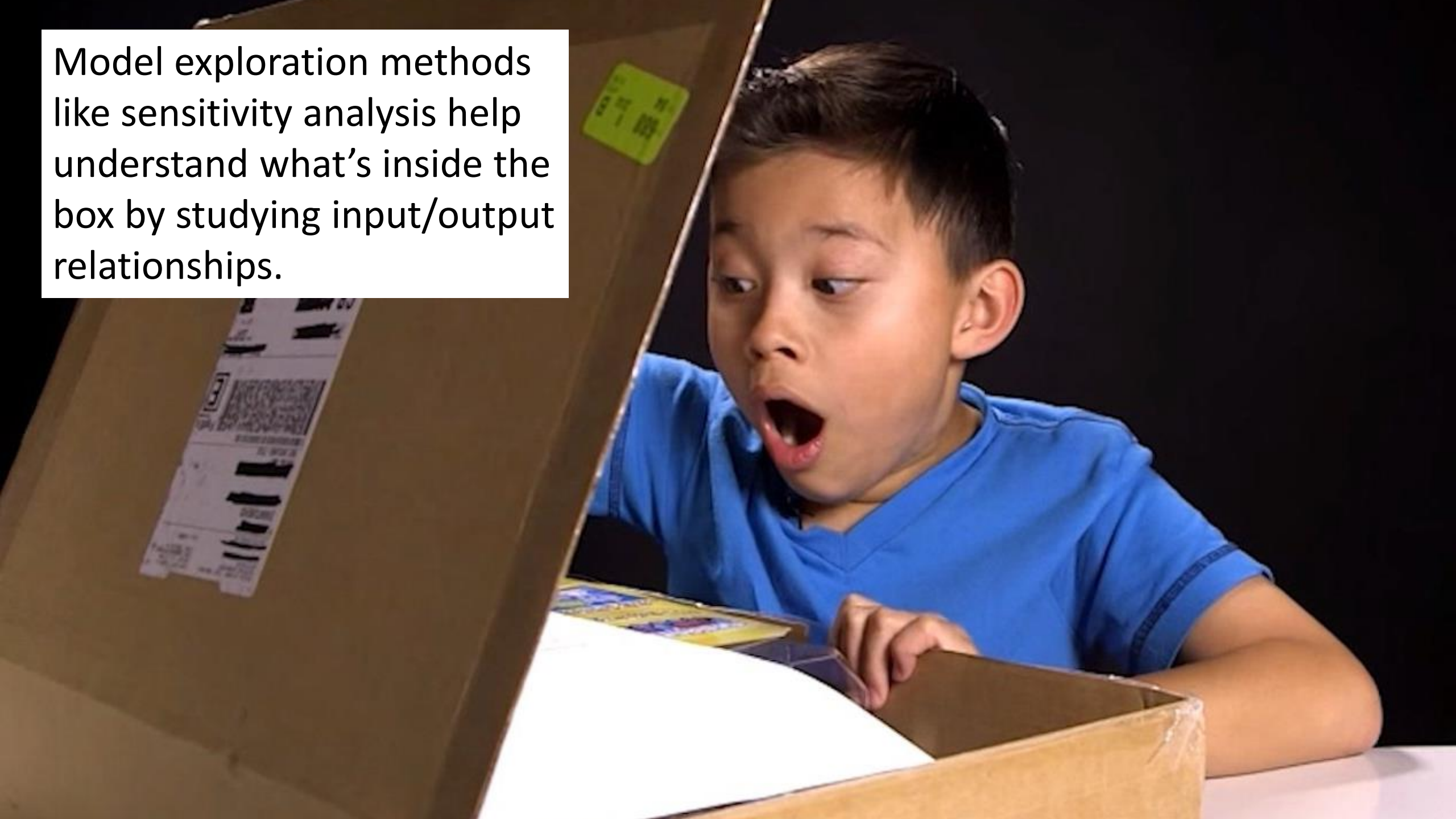
"Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad."

George Box, 1976.



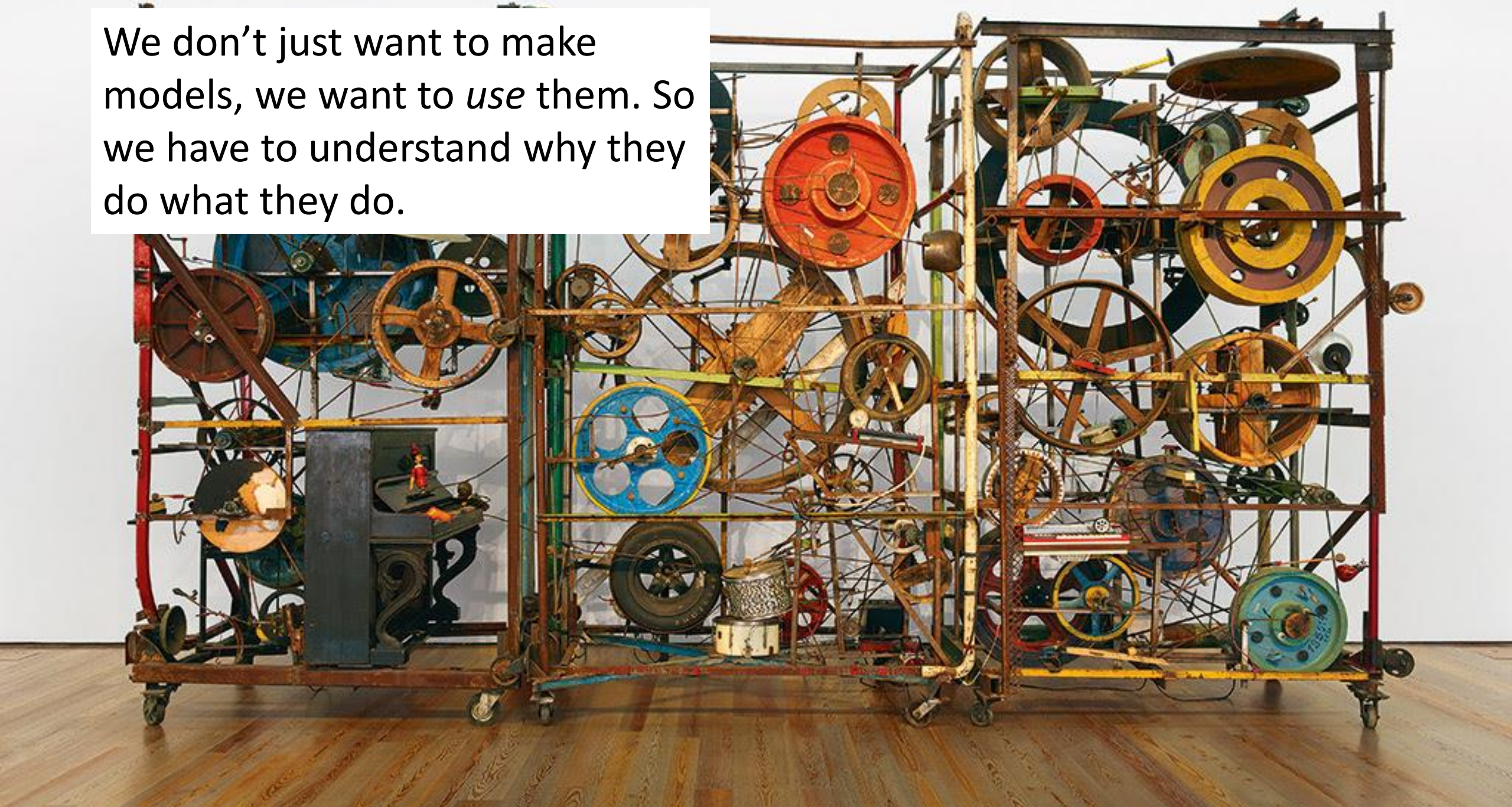
Social simulation models are black box functions. We put lots of data into them, and get some values out of them, but have no idea what happens inside.

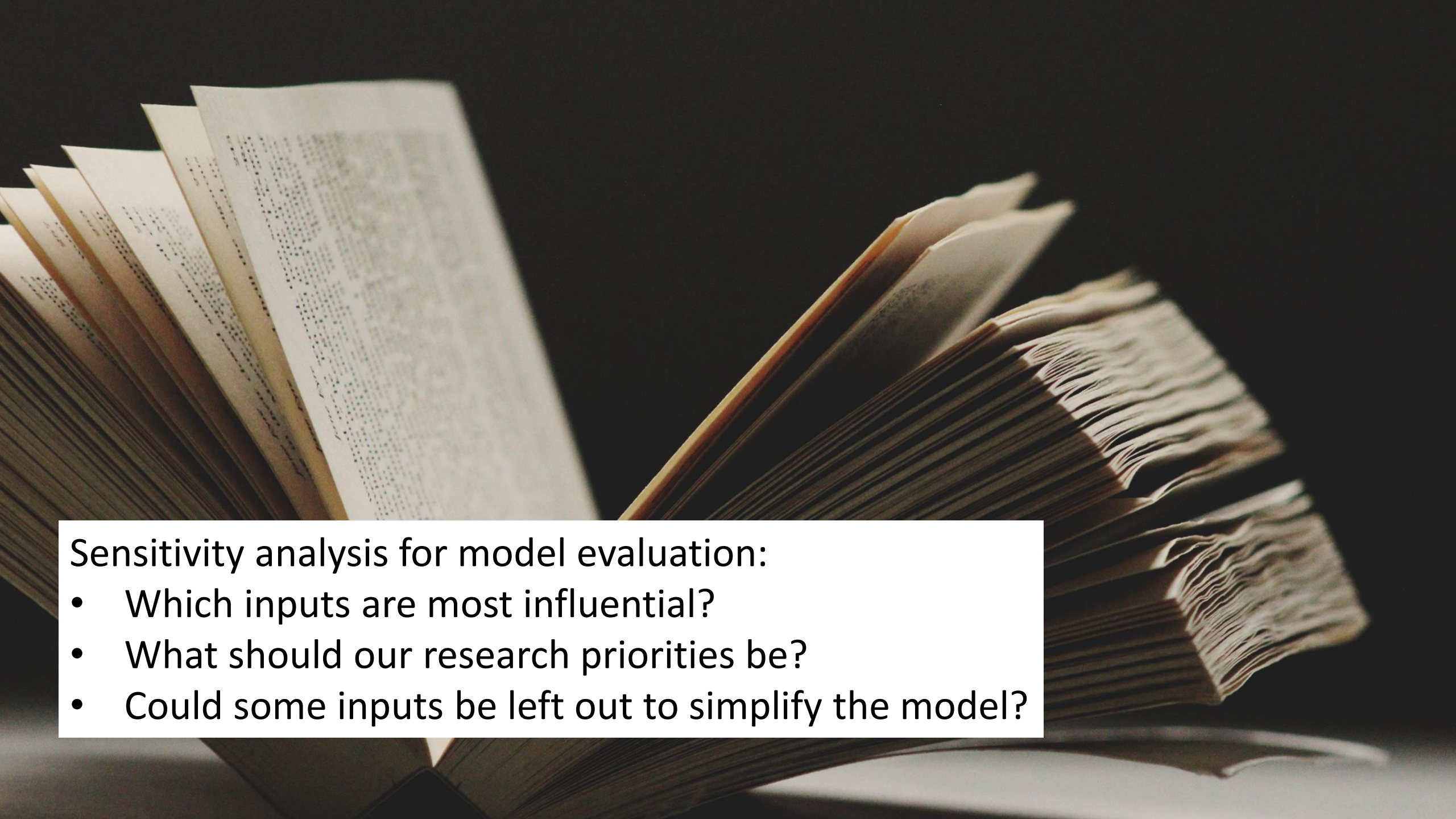
Model exploration methods like sensitivity analysis help understand what's inside the box by studying input/output relationships.



The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input (Saltelli et al., 2008)

We don’t just want to make models, we want to *use* them. So we have to understand why they do what they do.



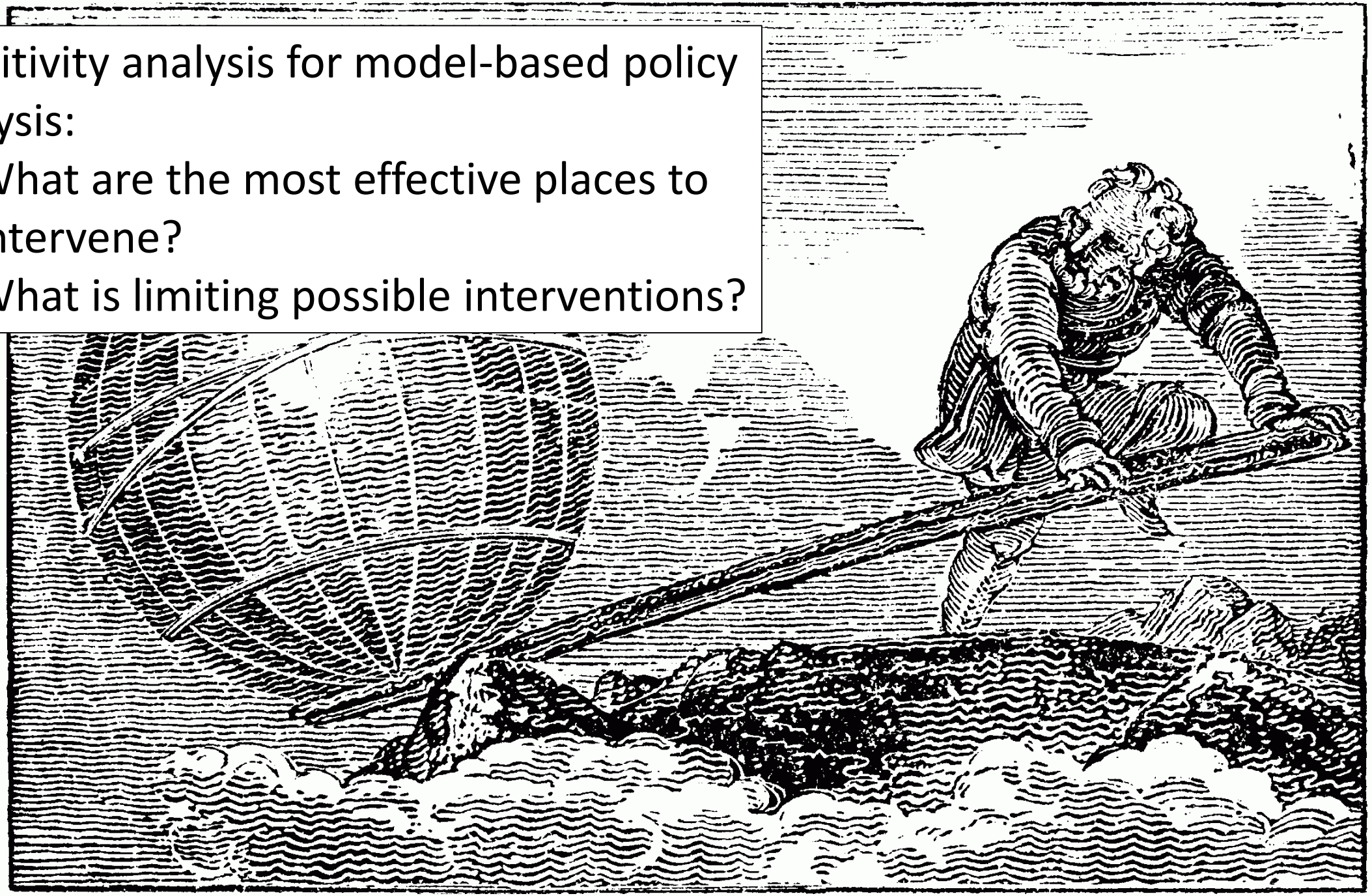


Sensitivity analysis for model evaluation:

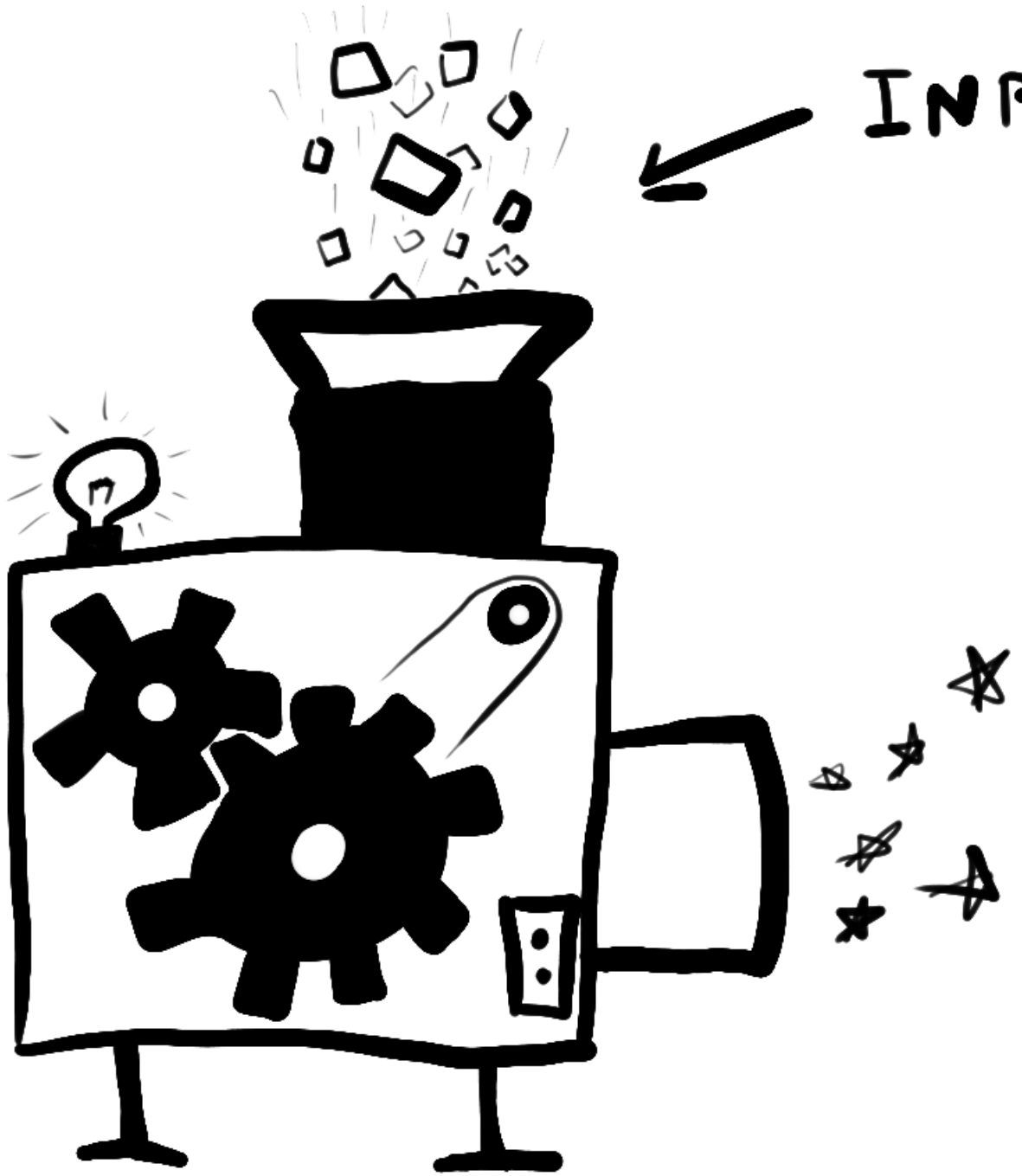
- Which inputs are most influential?
- What should our research priorities be?
- Could some inputs be left out to simplify the model?

Sensitivity analysis for model-based policy analysis:

- What are the most effective places to intervene?
- What is limiting possible interventions?




WE WANT YOU
TO ALWAYS DO
SENSITIVITY ANALYSIS
imgflip.com

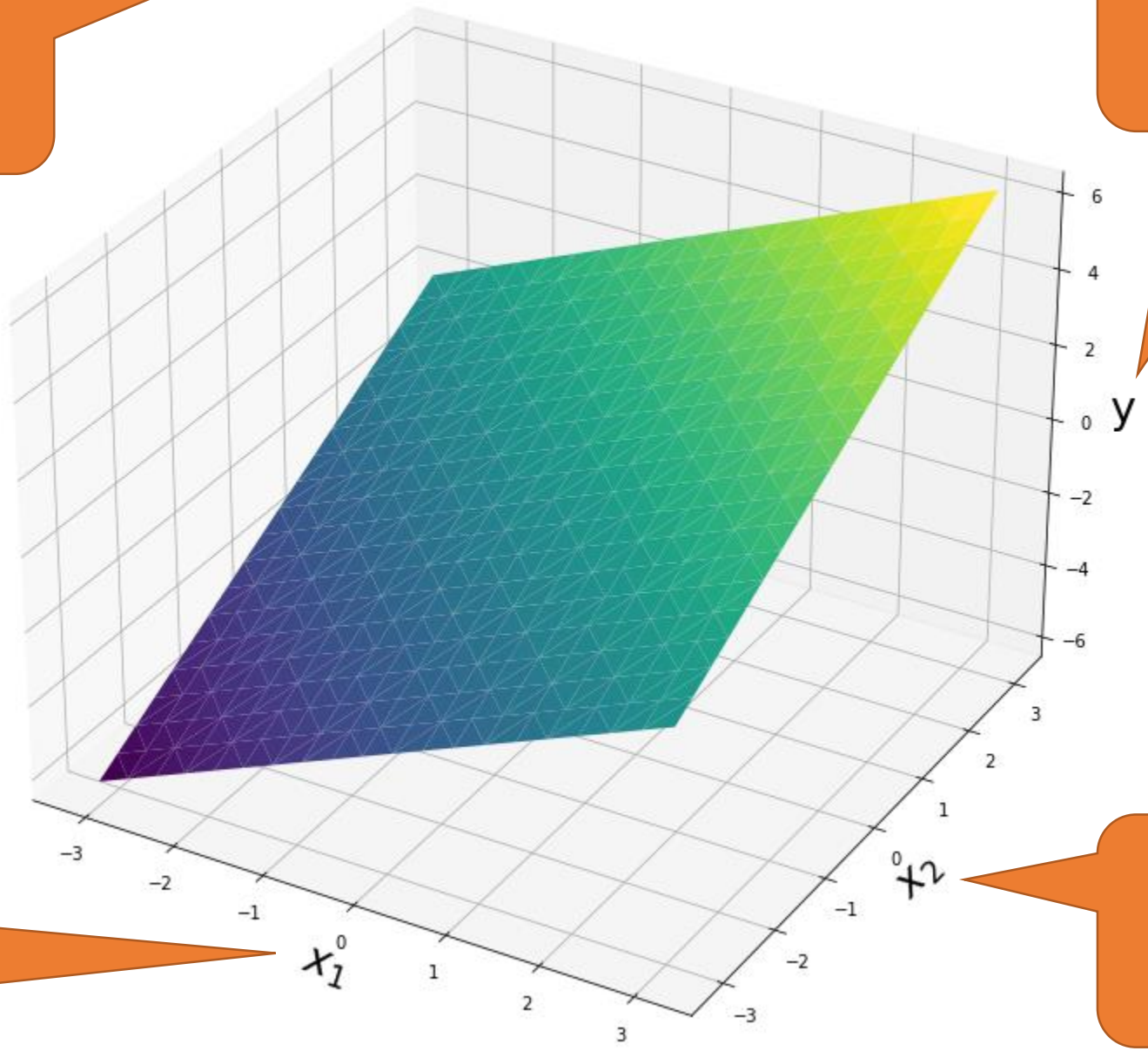


Let’s look at some basic examples of inputs and outputs, and how that can help us understand the function in between.

Model

$y = x_1 + x_2$

Output

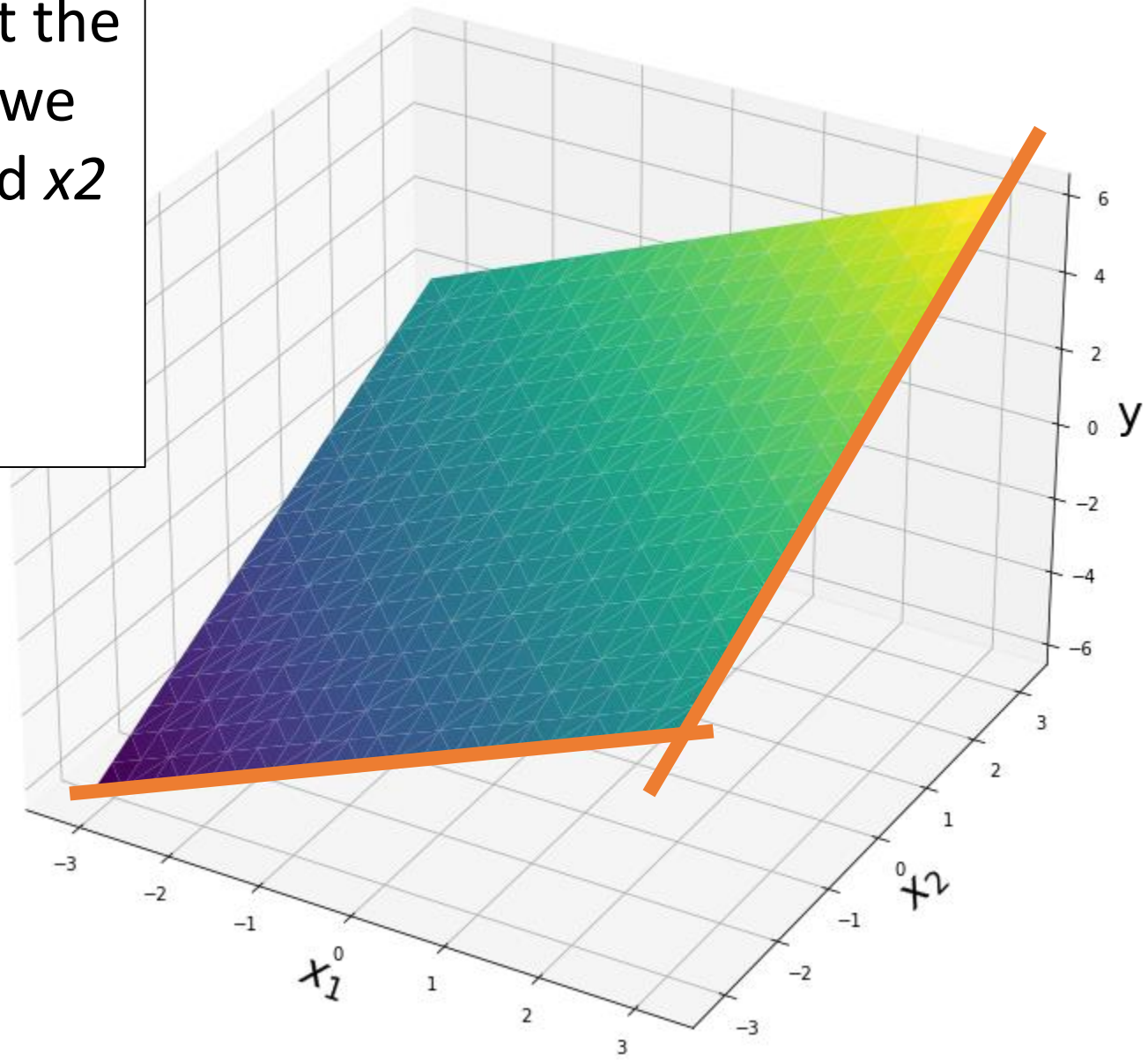


Input

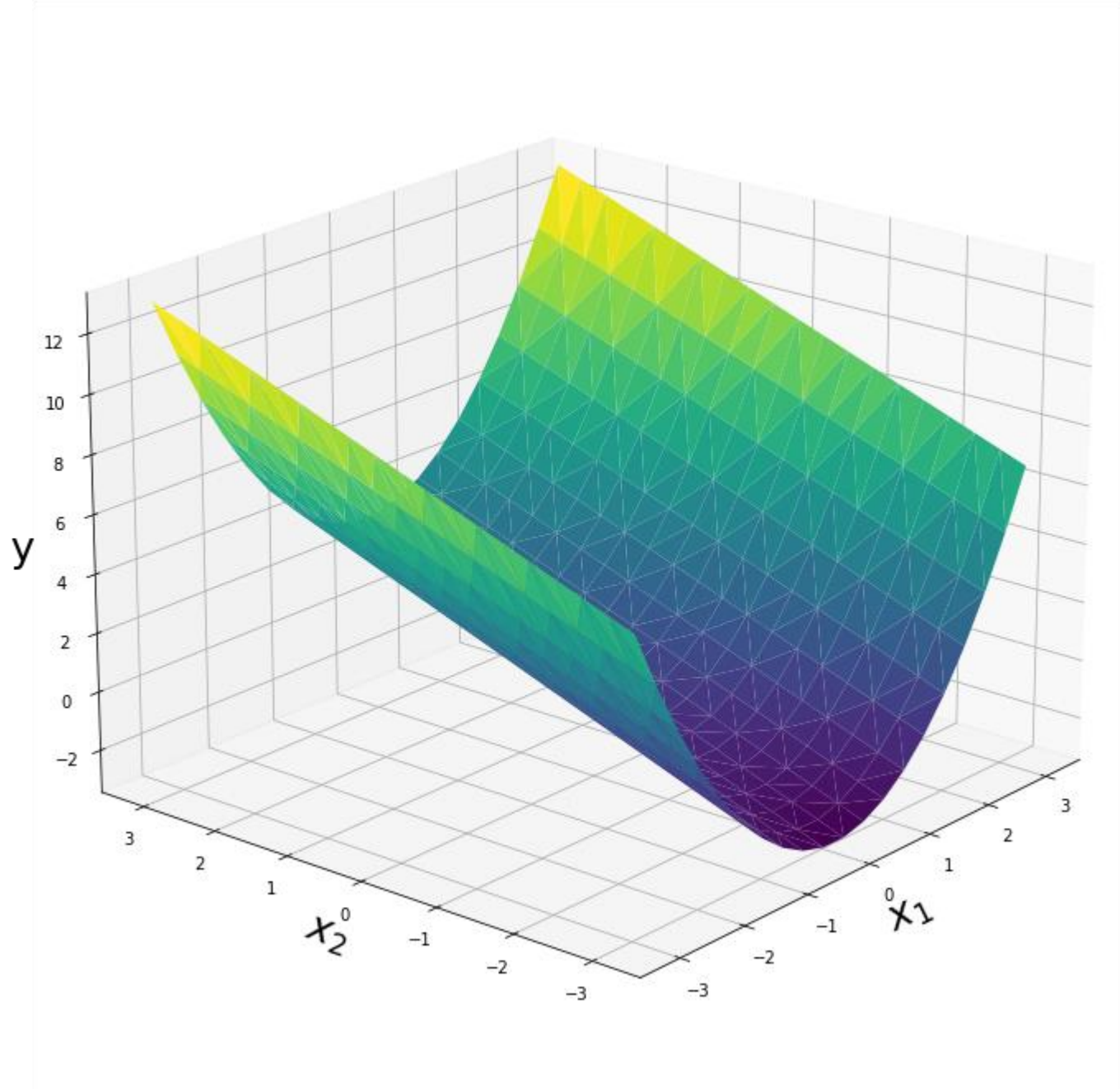
Input

$$y = x_1 + x_2$$

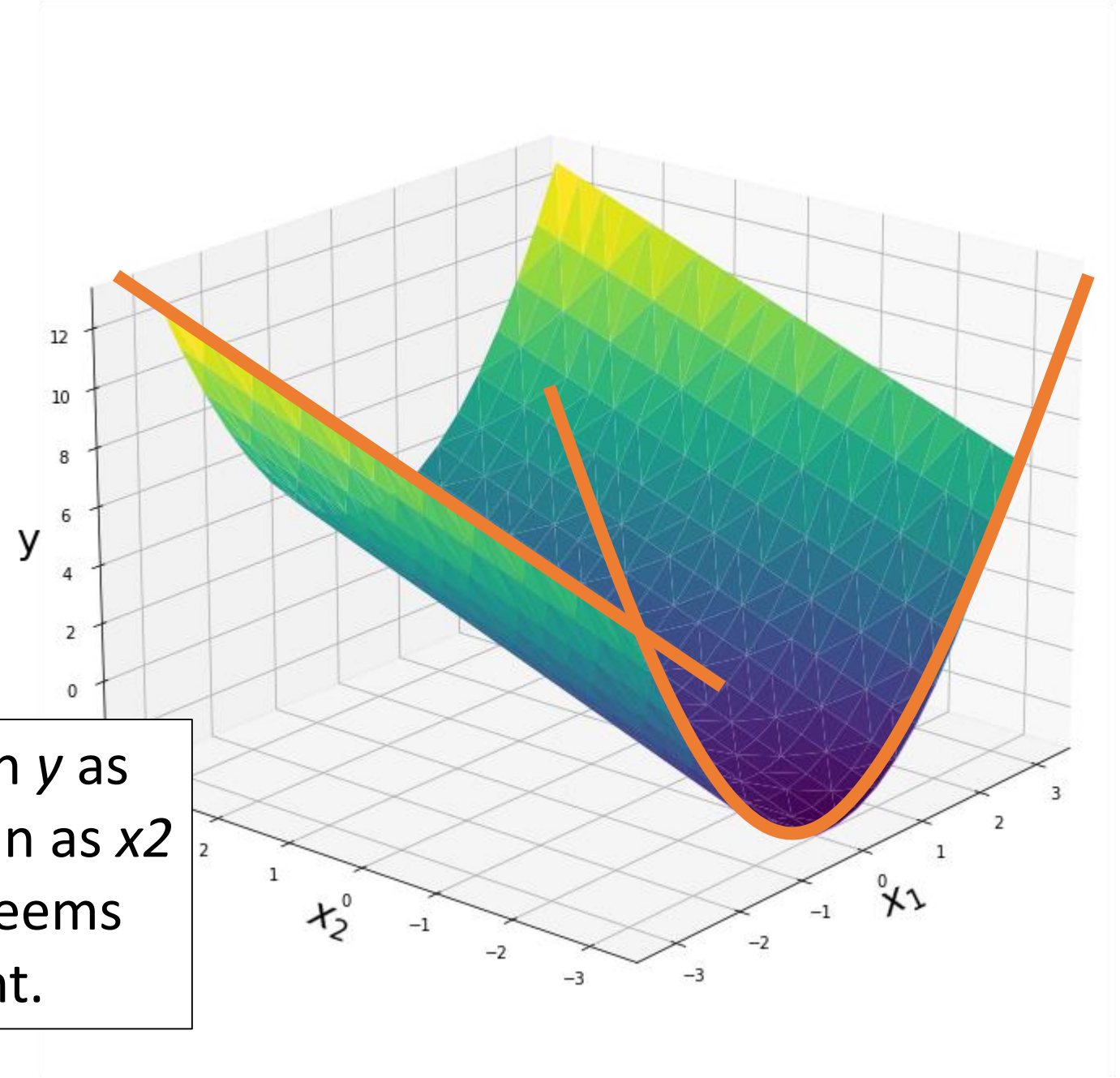
Without looking at the model (function), we can say that *x1* and *x2* are about equally important for the output.



$$y = x_1^2 + x_2$$



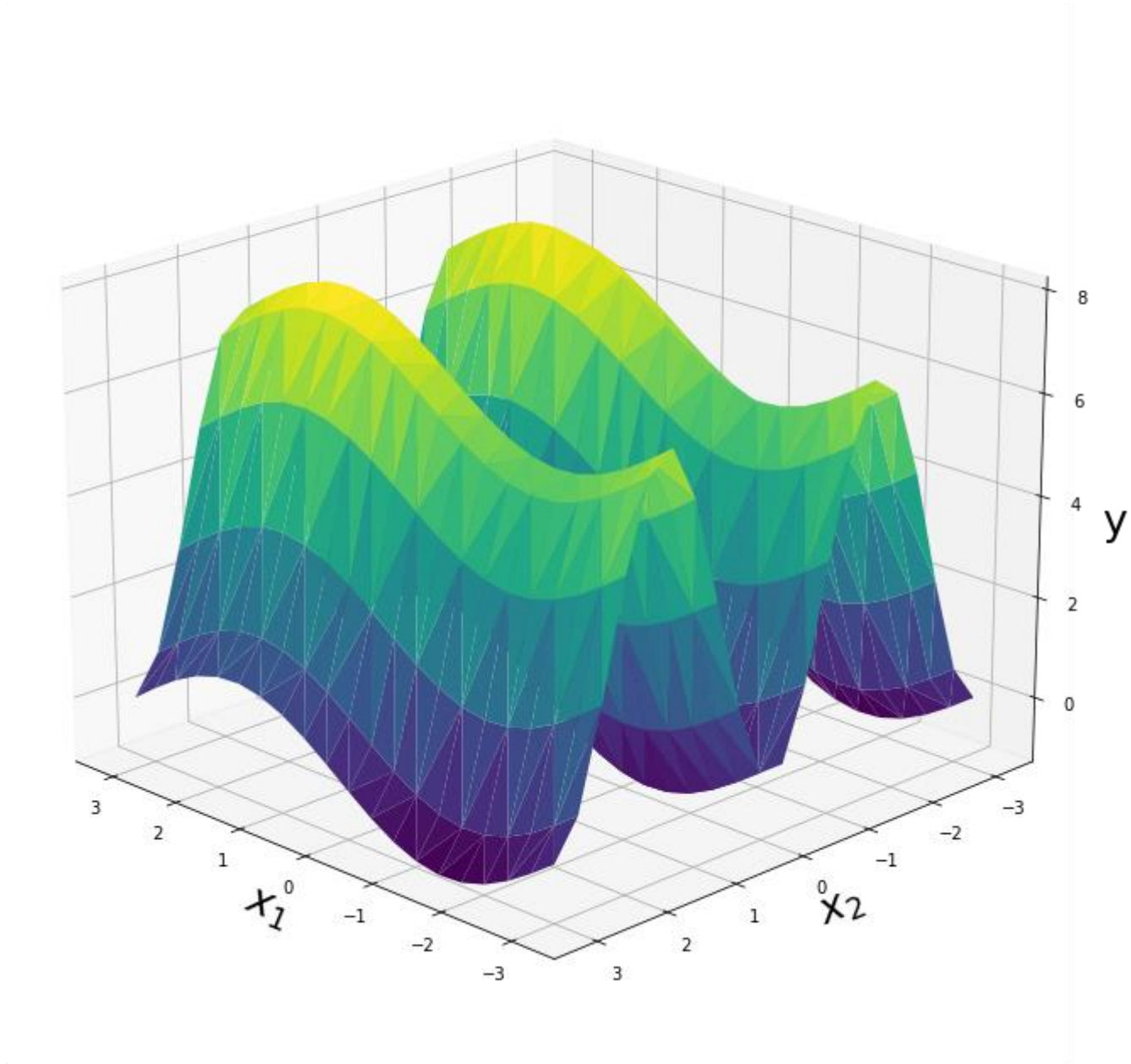
$$y = x_1^2 + x_2$$



More change in *y* as *x1* changes than as *x2* changes – *x1* seems more important.

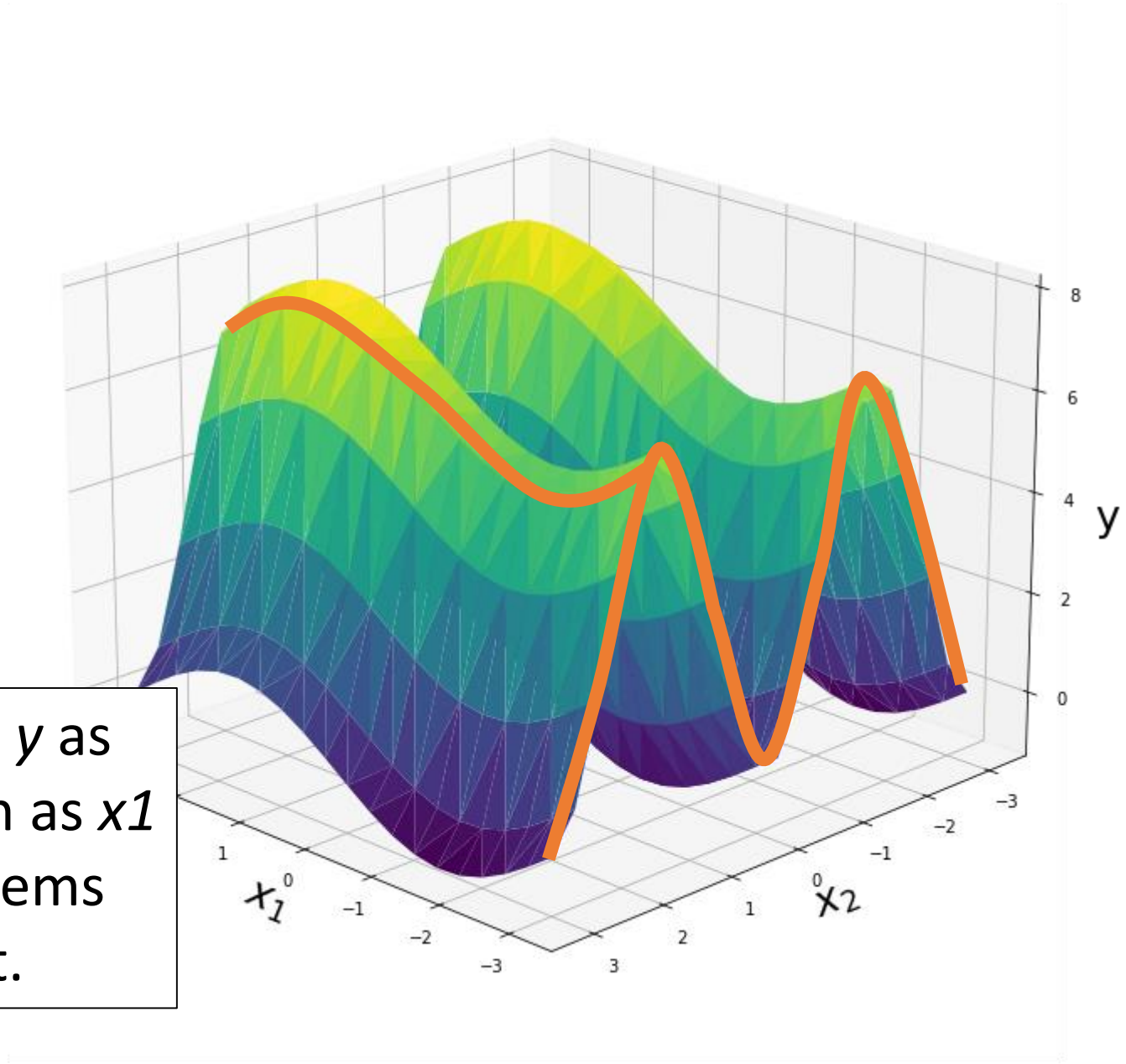
$$y = sin(x_1) + a * sin(x_2)^2 + b * x_3^4 * sin(x_1)$$

with $x_3 = 1$



$$y = sin(x_1) + a * sin(x_2)^2 + b * x_3^4 * sin(x_1)$$

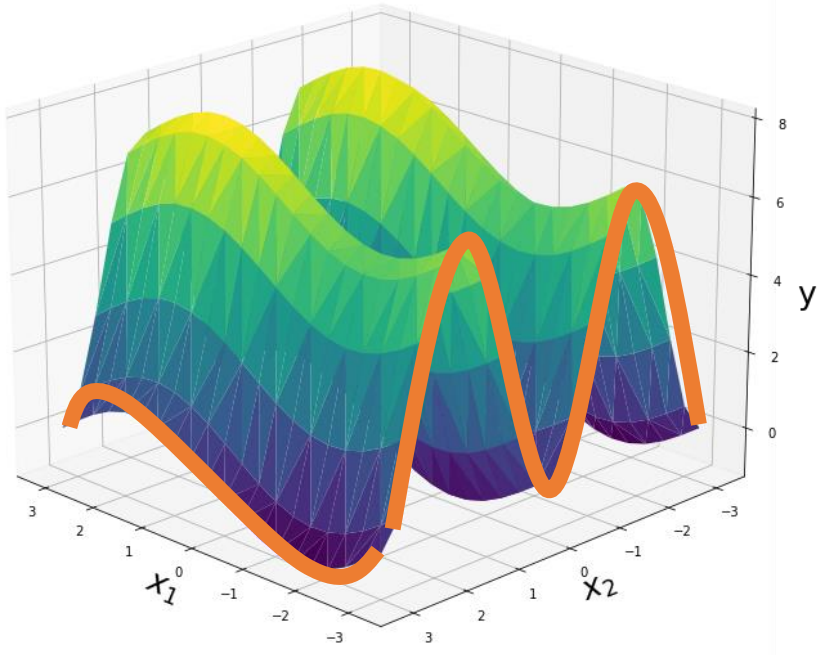
with $x_3 = 1$



More change in *y* as *x2* changes than as *x1* changes – *x2* seems more important.

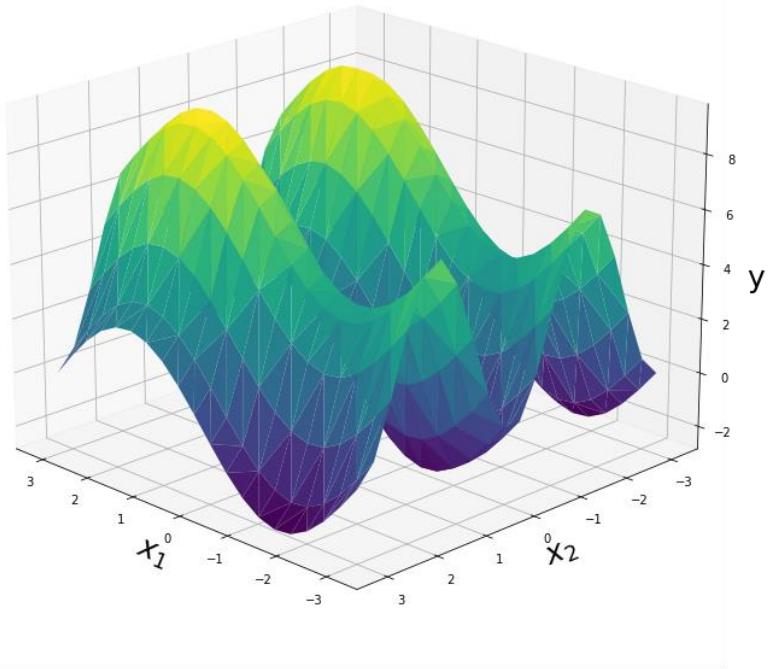
$y = sin(x_1) + a * sin(x_2)^2 + b * x_3^4 * sin(x_1)$

with $x_3 = 1$



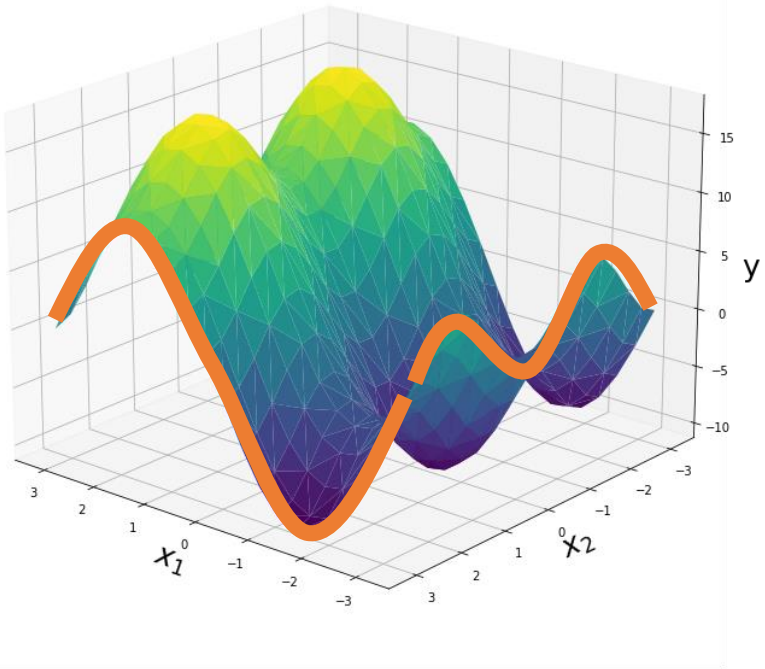
$y = sin(x_1) + a * sin(x_2)^2 + b * x_3^4 * sin(x_1)$

with $x_3 = 2$



$y = sin(x_1) + a * sin(x_2)^2 + b * x_3^4 * sin(x_1)$

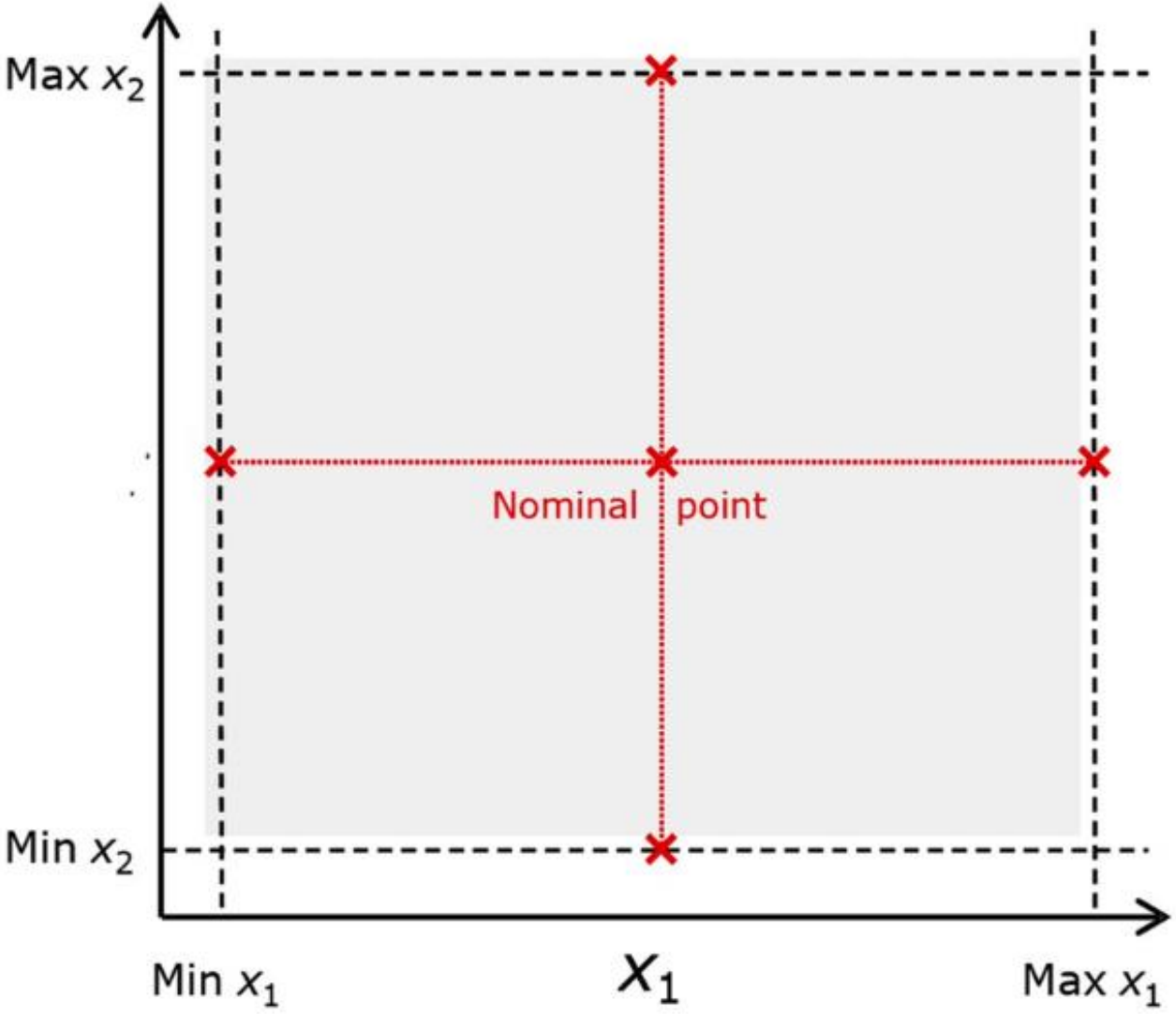
with $x_3 = \pi$

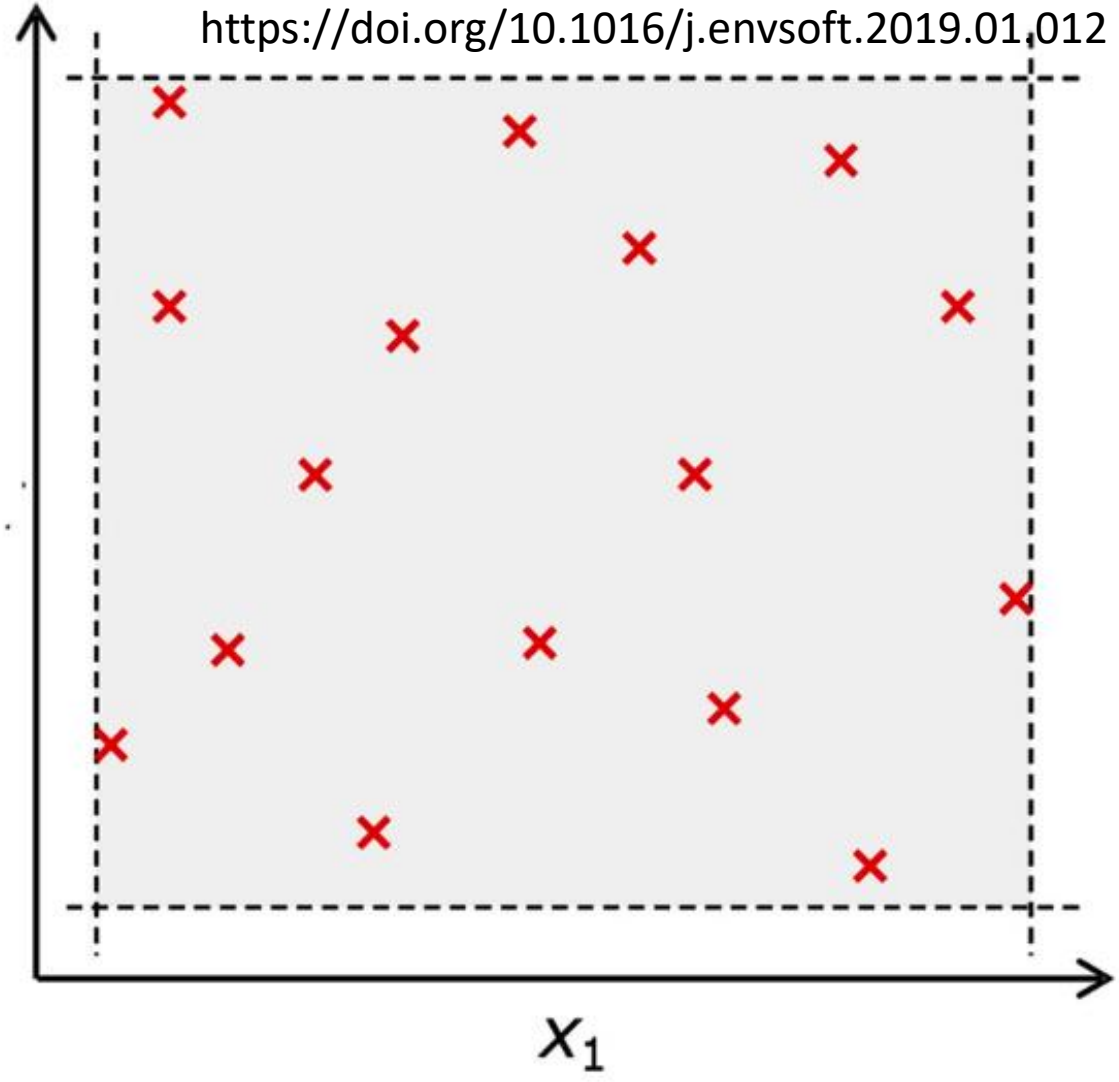


x_1 becomes more important than x_2 as x_3 changes!

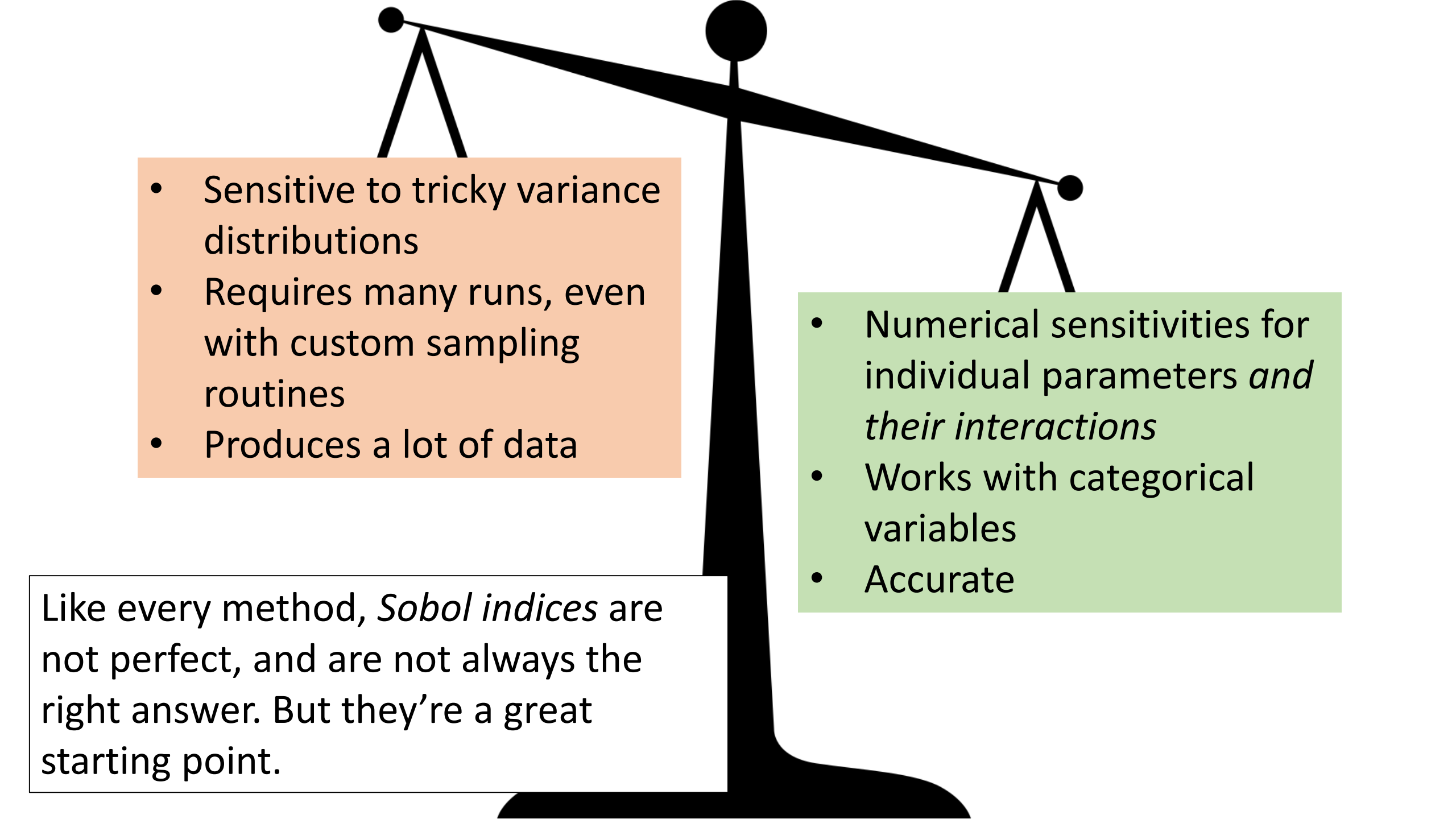

Many methods for measuring sensitivities exist. Gold standard: *Sobol indices*, a global sensitivity analysis method.
Strateline
Distinction
Grand Junction
SENECA FALLS, N.Y.
A GOOD RULE
"Do Unto Others As You Woul
PAT'D
PAT'S PEND

https://doi.org/10.1016/j.envsoft.2019.01.012





When in doubt, use global sensitivity analysis - local (or one-factor-at-a-time) only if you have very good reasons.

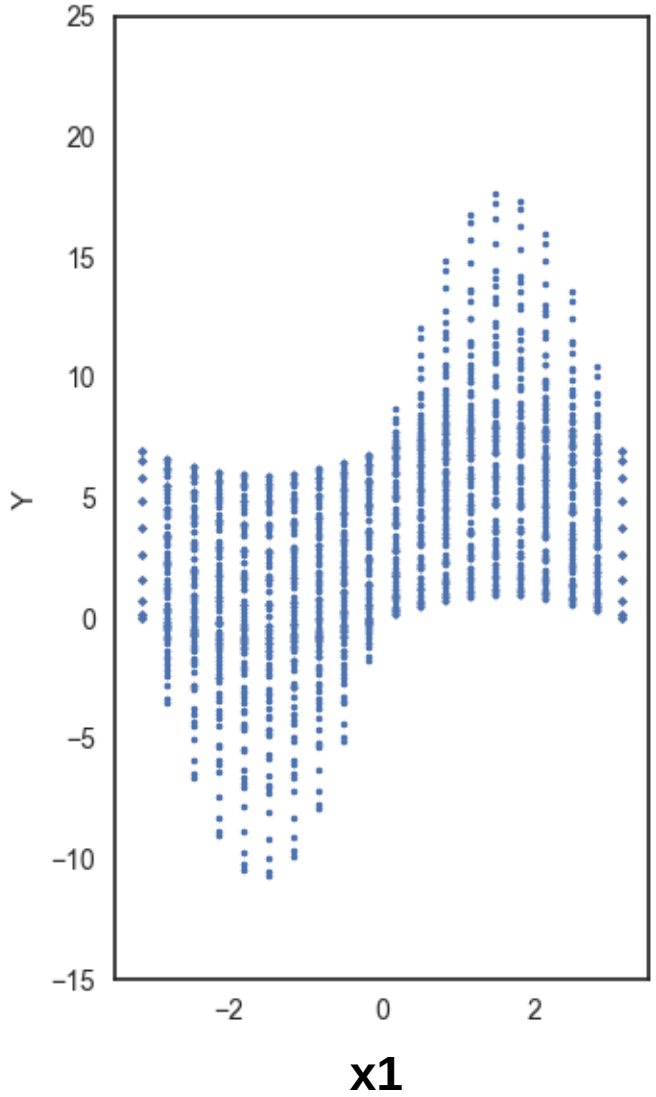


- Sensitive to tricky variance distributions
- Requires many runs, even with custom sampling routines
- Produces a lot of data

- Numerical sensitivities for individual parameters *and their interactions*
- Works with categorical variables
- Accurate

Like every method, *Sobol indices* are not perfect, and are not always the right answer. But they're a great starting point.

- Based on variance decomposition – tells us the fraction of total variance added by each variable

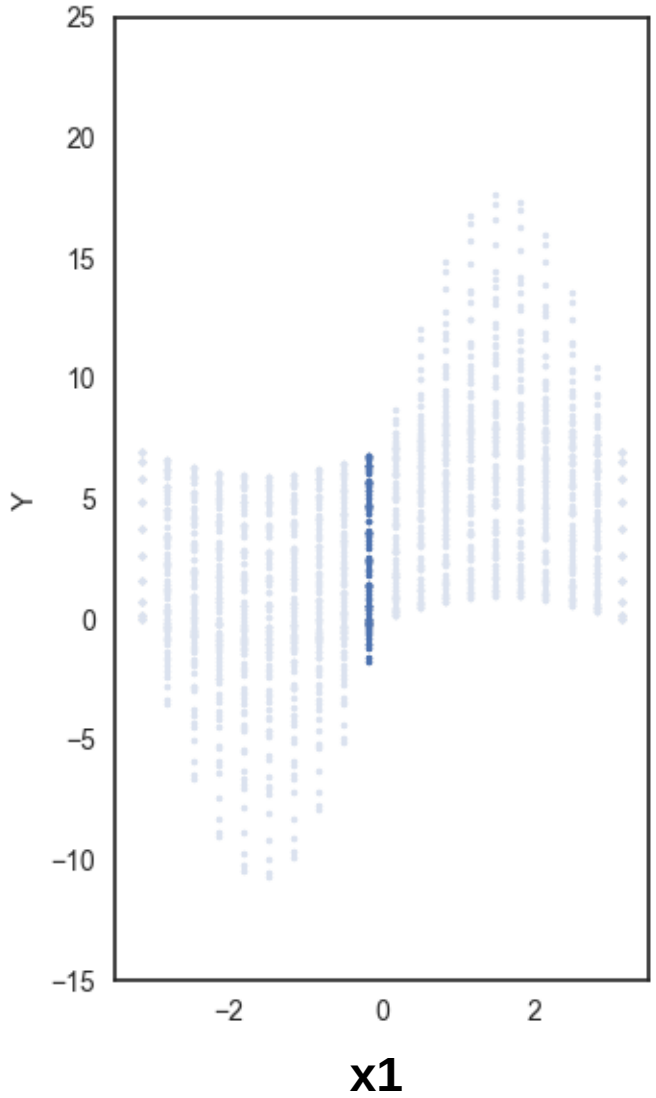


- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}\left[E_{x\sim x1}(Y|x1)\right] / V(Y)$$

V = variance
E = mean
$X\sim x_i$ = Set of inputs except x_i

- Based on variance decomposition – tells us the fraction of total variance added by each variable



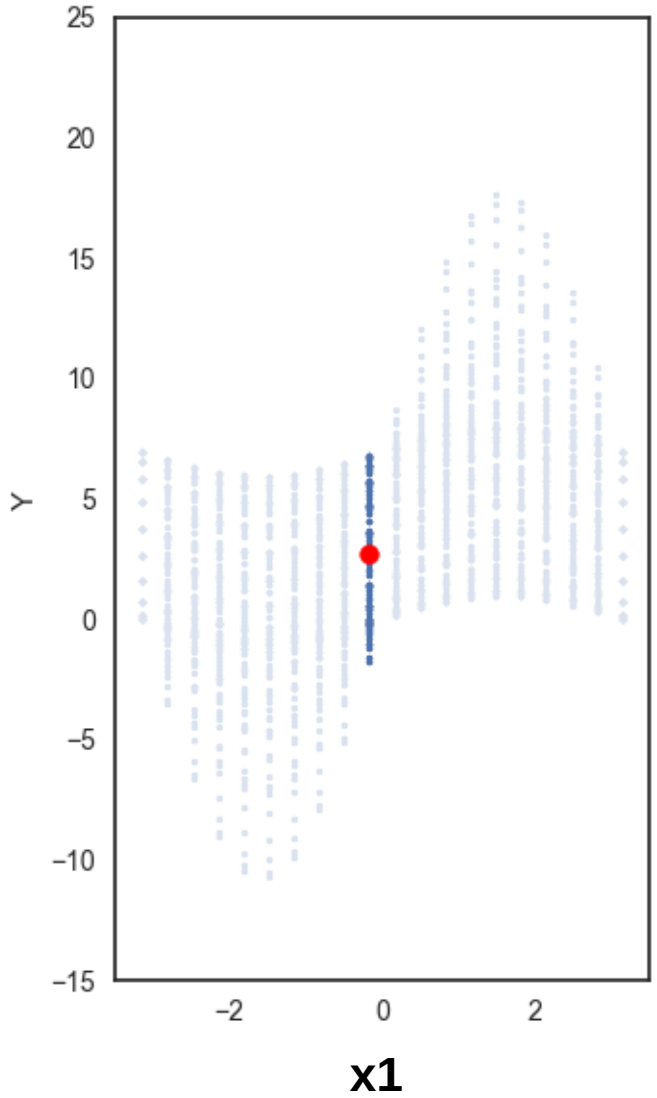
- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}[E_{x\sim x1}(Y|x1)] / V(Y)$$

x1 fixed
x2, x3 vary across their range

Slide by Marc Jaxa-Rozen

- Based on variance decomposition – tells us the fraction of total variance added by each variable

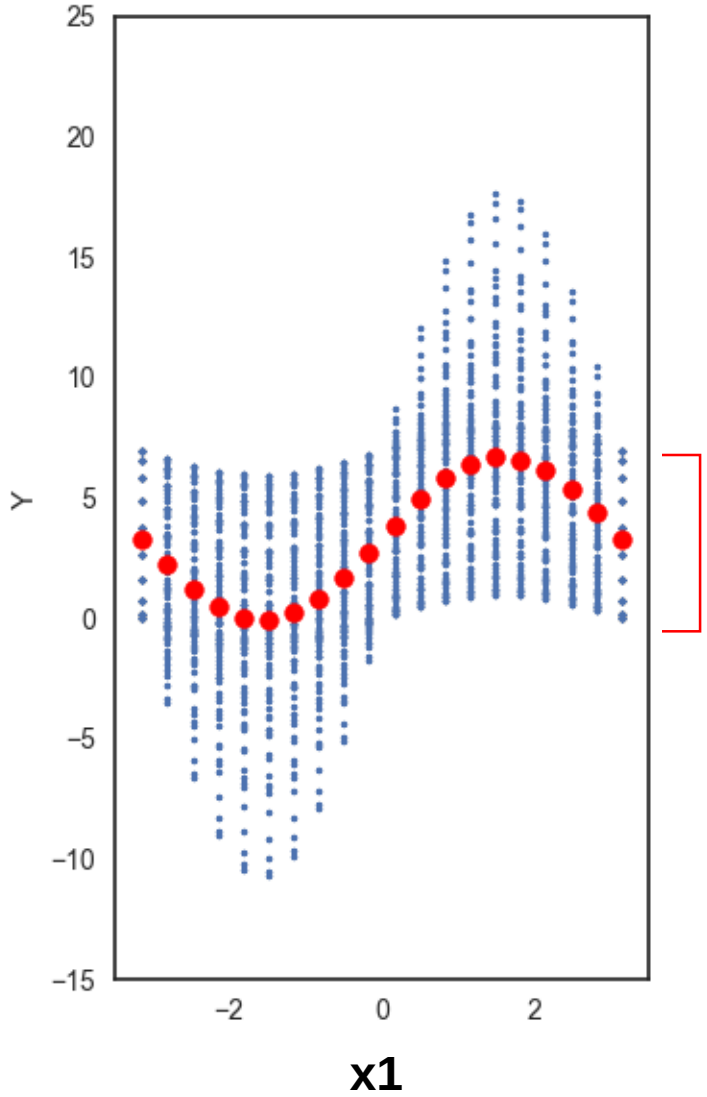


- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}[E_{x\sim x1}(Y|x1)] / V(Y)$$

Mean value across values of x2, x3

- Based on variance decomposition – tells us the fraction of total variance added by each variable

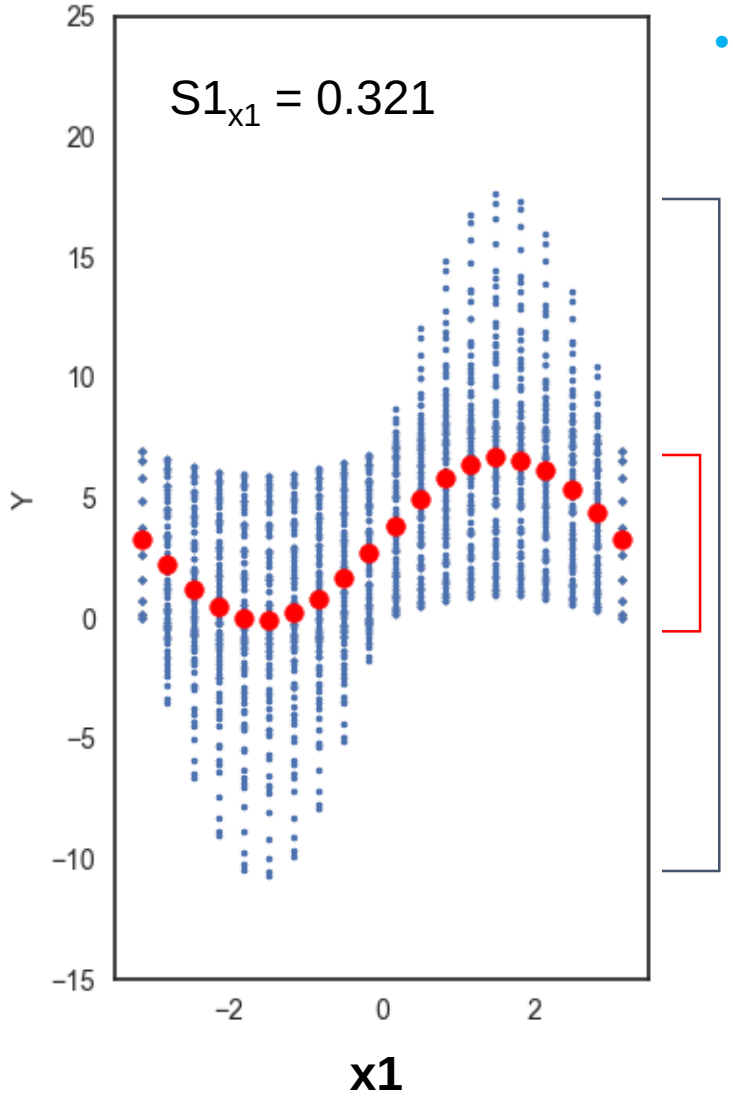


- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}\left[E_{x\sim x1}(Y|x1)\right] / V(Y)$$

Variance of the means across values of x1

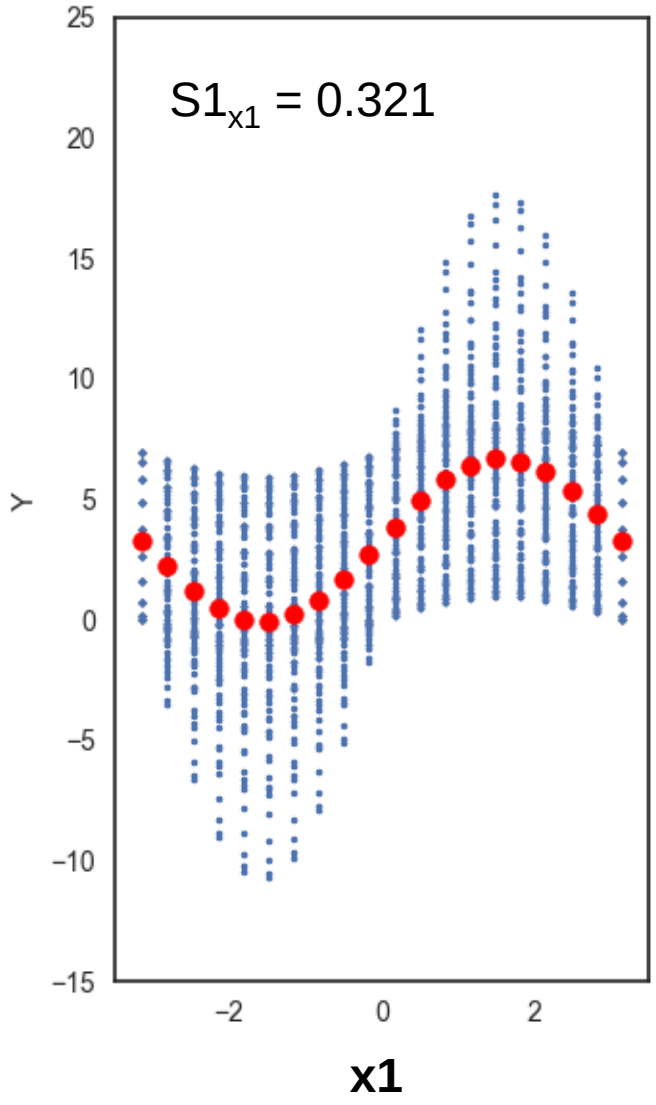
- Based on variance decomposition – tells us the fraction of total variance added by each variable



- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}\left[E_{x\sim x1}(Y|x1)\right] / V(Y)$$

- Based on variance decomposition – tells us the fraction of total variance added by each variable



- First-order effect (S1): e.g. how much does x1 add to the variance of Y on its own?

$$S1_{x1} = V_{x1}\,[E_{x\sim x1}(Y|x1)] \,/\, V(Y)$$

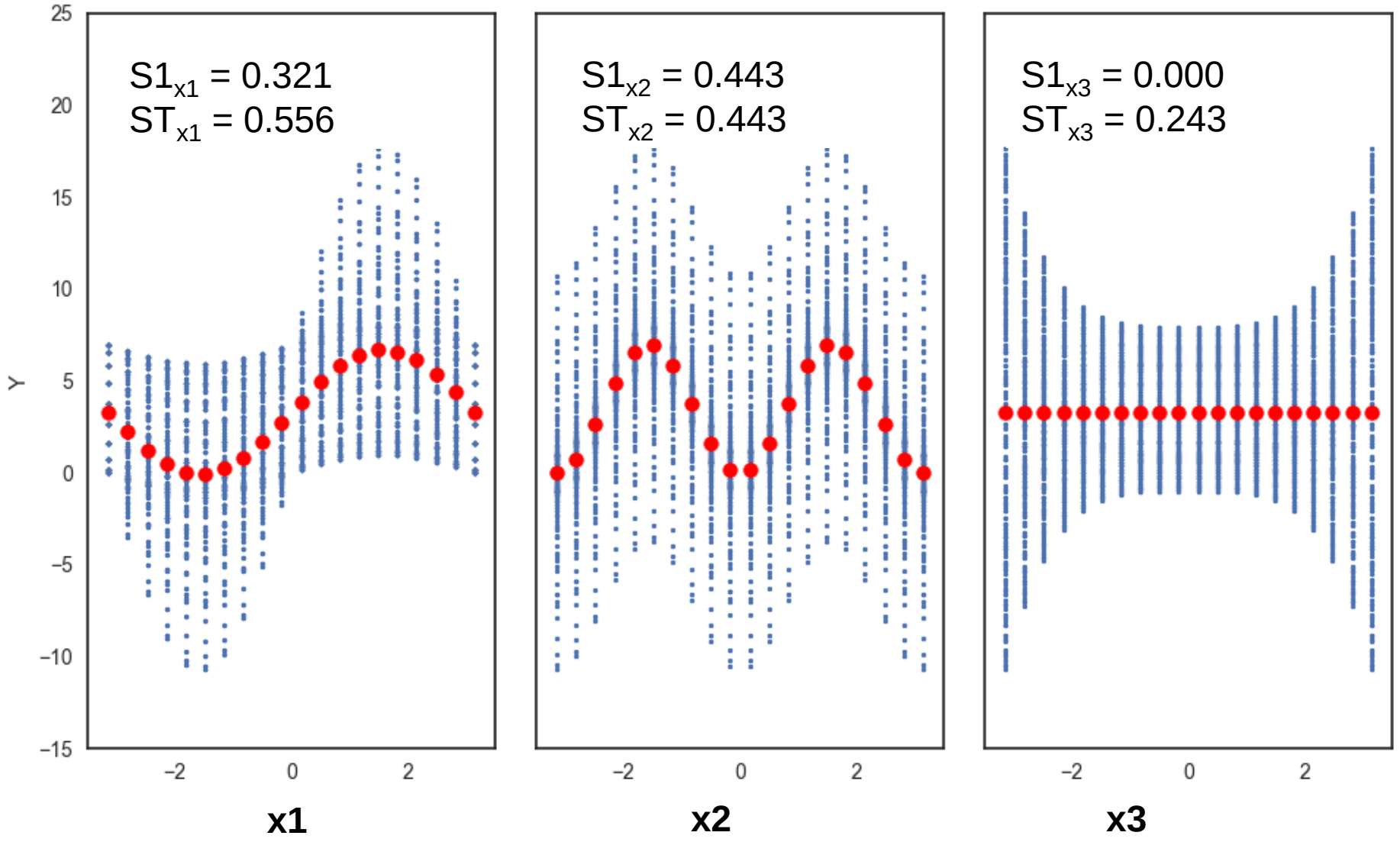
- Total effect (ST): e.g. how much does x1 add to the variance of Y, including all its interactions?

$$ST_{x1} = E_{x\sim x1}\,[V_{x1}(Y|X\sim x1)] \,/\, V(Y)$$

- Second-order effects (S2): e.g. how much specific interactions between x1 and x2 add to variance of Y

$$S2_{x1,x2} = E_{x\sim x1,x2}\,[V_{x1,x2}(Y|X\sim x1,x2)] \,/\, V(Y) - S1_{x1} - S1_{x2}$$

Slide by Marc Jaxa-Rozen



- For an additive model (no interactions):

  $ST = S1$

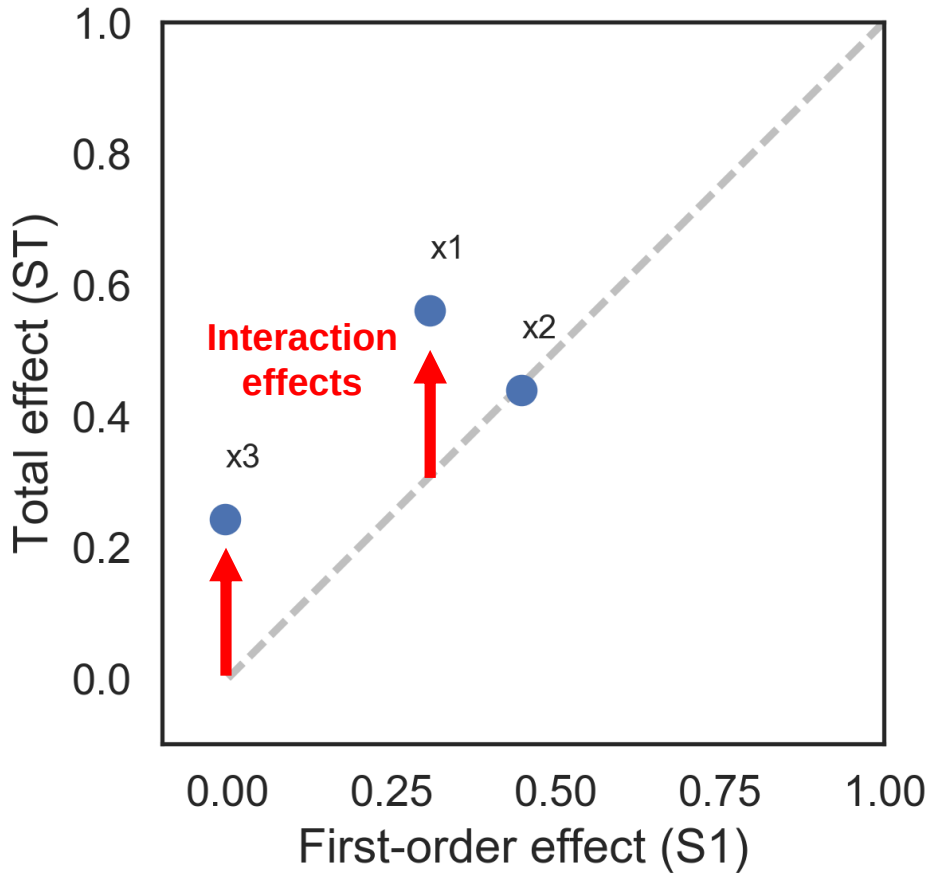
  $\sum S1 = 1$

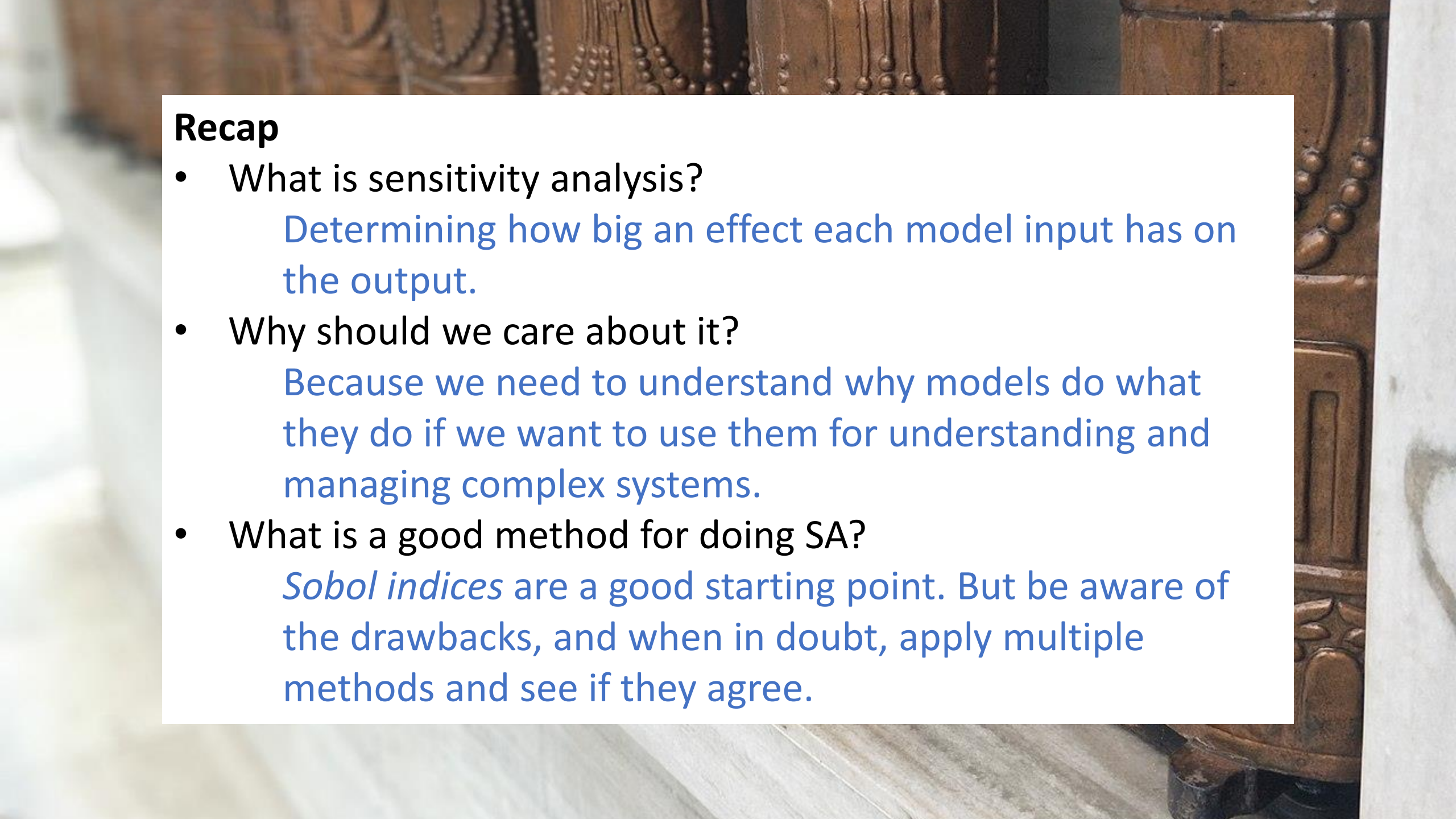
- Otherwise: $\sum S1 < 1$

  In our example: $\sum S1 = 0.764$

  So that 1 – 0.764 = 24% of output variance is caused by interactions between variables

- In general:
  - **Prioritize inputs with the highest S1 index**
  - **Discard inputs with the lowest ST index**





Recap

- What is sensitivity analysis?

  Determining how big an effect each model input has on the output.

- Why should we care about it?

  Because we need to understand why models do what they do if we want to use them for understanding and managing complex systems.

- What is a good method for doing SA?

  Sobol indices are a good starting point. But be aware of the drawbacks, and when in doubt, apply multiple methods and see if they agree.