Model predictions

We provide more detail here on how the different accounts shown in Figure 2 were fit to the data.

1 Models

1.1 Abnormal selection

The abnormal selection model predicts that abnormal causes will be selected (Halpern & Hitchcock, 2015; Hart & Honoré, 1959/1985; Hilton & Slugoski, 1986; Hitchcock & Knobe, 2009). It predicts that participants will prefer the ball that normally gets blocked when both balls go through, and the ball that normally goes through when both balls get blocked. We encoded this predicted preference by setting the abnormal cause to 1, and the normal cause to 0.

1.2 Correspondence

The correspondence model predicts that abnormal causes are selected for abnormal effects, and normal causes for normal effects. To make the predictions of this model precise for our scenarios, we compared the probability of each candidate causal event (e.g. ball A going through the gate) with the probability of the effect event (e.g. ball E going through the gate). We defined the correspondence between cause C and effect E as

$$\text{correspondence}(C, E) = 1 - |p(C) - p(E)|, \tag{1}$$

where $p(C)$ and $p(E)$ are the probabilities of observing this particular cause event and effect event, respectively.

1.3 Necessity and sufficiency

On one way of making the necessity and sufficiency view precise (Icard, Kominsky, & Knobe, 2017), we can define the causal strength of a cause C for an effect E, written $\kappa(C, E)$, as:

$$\kappa(C, E) = p(C = 0) \cdot p(E = 0 \mid A = 1, do(C = 0)) + p(C = 1) \cdot p(E = 1 \mid do(C = 1)). \tag{2}$$

In words, this equation is the weighted sum of the "actual necessity" of C for E (given by the probability that E would have been 0 upon an intervention setting C to 0, given that in fact the alternative cause A was 1), and the "robust sufficiency" of C for E (given by the probability that E would have been 1 upon an intervention setting C to 1). The weighting is determined by the probability of the cause (see Icard et al., 2017, for details and discussion).

2 Model fitting

Some of the models discussed above yield a continuous measure for each candidate cause. In our experiment, we asked participants to choose between the two candidate causes. To model participants' choices, we used a softmax function that maps from the two continuous measures onto a single value which expresses the probability of selecting cause i from a set of potential causes $\mathcal{C}$ as

$$\text{p(select cause}_\text{i}) = \frac{exp(\text{cause}_\text{i} \cdot \beta)}{\sum\limits_{j \in \mathcal{C}} exp(\text{cause}_\text{j} \cdot \beta)}, \tag{3}$$

where the measure for each cause is specified by the different models discussed above. β is a free parameter used to fit each model's predictions to participants' responses. We fit the models by minimizing the squared error between the predicted and observed probability of selecting one cause over the other. The best-fitting β parameter for the *abnormal selection model*, the *correspondence model*, and the *necessity and sufficiency model* were 1, 3.2, and 2.7, respectively.

References

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *British Journal for the Philosophy of Science*, *66*, 413–457.

Hart, H. L. A., & Honoré, T. (1959/1985). *Causation in the law.* New York: Oxford University Press.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*(1), 75–88.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, *11*, 587–612.

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93. Retrieved from `https://doi.org/10.1016%2Fj.cognition.2017.01.010` doi: 10.1016/j.cognition.2017.01.010