A Normative calculations

Normative causal attribution involves three steps: 1) attributing causes to effects that have occurred; 2) explaining away effects that should or might have occurred but were not observed; 3) examining the temporal distance between presumed preventative events and the subsequent effect event. The Step 1 and 2 correspond to path construction in the main text. We use $\{\alpha_g, \beta_g\}$,$\{\alpha_p, \beta_p\}$, $\{\alpha_b, \beta_b\}$ to denote parameters of gamma distributions for generative delays, preventative windows, and base rate delays.

Step 1 is to form $g' \rightarrow e'$ pairs where 1) the effect event e' is not over-determined (i.e. has a single actual cause), 2) the cause event g' does not produce its effect twice, and 3) g' precedes e'. The likelihood of each pair is then determined by mapping the delay between g' and e' to the gamma density function:

$$P(g' \rightarrow e'|\alpha_g, \beta_g) = P(t_{g' \rightarrow e'} = t_{g'e'}|\alpha_g, \beta_g) \tag{A.1}$$

Step 2 involves forming $g' \rightarrow h$ pairs where h is a hidden effect event assumed to happen some time after the observable period *or* at some point during a preventative window. The likelihood calculation depends on the gamma cumulative density falling beyond the end of the clip or within the window:

$$\begin{aligned} P(g' \rightarrow h|\alpha_g, \beta_g, \alpha_p, \beta_p) = P(t_{g' \rightarrow h} > t_{end}|\alpha_g, \beta_g) + \\ P(t_{g' \rightarrow h} t_{end}|\alpha_g, \beta_g) \prod_{p'} P(t_{g' \rightarrow h} < t_{g'h} + t_{p' \rightarrow h}|\alpha_g, \beta_g, \alpha_p, \beta_p) \end{aligned} \tag{A.2}$$

Base rate activations of the effect event are represented as having been caused by the previous base rate activation, which can also be represented as $g' \rightarrow e'$ pairs where g' is actually the target component's (i.e., E) activation. When there are presumed preventative cause events, the base rate activation could be prevented but then subsequently "recover". Therefore, for base rate activation we could jointly consider Step 1 and Step 2 as $g' \rightarrow h^{(1)} \rightarrow \ldots \rightarrow h^{(n)} \rightarrow e'$, where $h^{(1)} \ldots h^{(n)}$ happens within the preventative windows. Meanwhile, according to the summing property the gamma distribution, if $X, Y \sim Gamma(\alpha, \beta)$ then $X + Y \sim Gamma(2\alpha, \beta)$. The probability $P(g' \rightarrow h^{(1)} \rightarrow \ldots \rightarrow h^{(n)} \rightarrow e')$ can thus be represented as Eq. A.3, where the calculation of $P(g' \rightarrow e')$ is similar to Eq. A.1, and the calculation of $P(g' \rightarrow h^{(n')})$ is similar to Eq. A.2 except that t_{end} is substituted with $t_{e'}$ and only the second item of prevention is considered.

$$\begin{aligned} P(g' \rightarrow h^{(1)} \rightarrow \ldots \rightarrow h^{(n)} \rightarrow e'|\alpha_b, \beta_b, \alpha_p, \beta_p) = \\ P(g' \rightarrow e'|(n+1) \cdot \alpha_b, \beta_b) \prod_{n' \in n} P(g' \rightarrow h^{(n')}|n \cdot \alpha_b, \beta_b, \alpha_p, \beta_p) \end{aligned} \tag{A.3}$$

Finally, the prevention examination in Step 3 extracts all presumed preventative events and their nearest effect events to form $p' \rightarrow e'$ pairs (there is no need for examination if no effect events happen after p'), and then applies gamma cumulative density function of prevention:

$$P(p' \rightarrow e'|\alpha_p, \beta_p) = P(t_{p' \rightarrow e'} < t_{p'e'}|\alpha_p, \beta_p) \tag{A.4}$$

B Simulation-and-summary calculations

Characteristic summary statistics for each structure hypothesis were constructed by simulating 10,000 sequences of point events from each structure type, with three interventions on A or B, and then calculating the empirical features for each intervention in each structure. This results in 60,000 simulated cases. Distinct from the experimental stimuli, simulated sequences here were not cut at twenty seconds so as to avoid the complex boundary effect in distribution constructions. By its definition we can see that the delay cue is independent of segmentation approaches since it always relates to the nearest effect event, while the count cue is sensitive on the segmentation for which we need to build distributions for intervention-based and fixed-window assumptions separately. Delay distributions use the probability density function smoothed with Gaussian kernels, and Count distributions used the discrete probability mass functions directly. When observing a new interventions, the probability of each causal structure was estimated by the normalized posterior of the summary statistic calculated on the observed data.

Inherent to this heuristic approach is the radical simplifying assumption that the features of the evidence subsequent to each control component event are modular and independent, that is, that

one can safely ignore that the subsequent device behavior also depends on the behavior of the other control component(s). Thus, each connection was estimated independently as generative, non-causal, or preventative, and then combined to yield a probability for each causal structure. For example, an intervention on A with a nearest effect occurring 2.5 seconds later has a posterior of [.2, .7, .1] of having being produced by a generative, non-causal or preventative $A \rightarrow E$ connection respectively under the regular base rate and [.3, .6, .2] under the irregular base rate (under the assumption of uniform prior distributions). When the next intervention on A happens, the likelihood will be updated by combining the new probability with the original one.

The boundary situations we considered were as follows: If no effect happens within the observation window, in both segmentation approaches, the delay cue will be marked as larger than the observing window and the probability will be estimated according to cumulative density function. If the observation window is less than the designed window length in the fixed-window approach (which often happens near the end of the clip), or there is no next intervention in the intervention-based approach, the count cue will be marked as greater than or equal to the observed count of effects and the probability will also be estimated on the basis of cumulative mass functions.

C Experiment stimuli generation and allocation

To ensure participants' performance on different conditions were comparable, the stimuli generation and assignment procedure was as follows: In Experiment 1, eighteen seeds were created independently. Each of them included a set of timings of interventions, regular base rate activations, irregular base rate activations, and what generative delays (or blocking windows) A and B would have if they were generative (or preventative) components. Then under each seed, 18 stimuli (9 causal structures $\times$ 2 base rate settings) were generated by implementing generative or preventative influences according to the grounded structure. All stimuli were finally divided into 18 sets (9 sets for each base rate setting) according to the Latin-square design that ensured participants would only see only one structure under each seed. Participants were randomly assigned to one of 18 sets. The half of the stimuli in Experiment 2 that have ground-truth answers also followed the procedure above.

D Softmax rules

We assumed that participants selected their response according to a softmax over a posterior value vector v:

$$P(n) = \frac{\exp(v_n/\tau)}{\sum_{n' \in N} \exp(v_{n'}/\tau)} \tag{D.1}$$

The "temperature" parameter $\tau \in (0, +\infty]$ controls how consistent the participant is in selecting the answer with the largest v_n in choice n. Smaller τ means that the participant's answer is better aligned with the model's answer with τ approaching $+\infty$ modeling random selection. For the normative model we simply set v_n to $P(s|\mathbf{d}, \mathbf{w})_n$, as well as the single cue models in the stimulation-and-summary approach. For the combination of two cues, we use two temperatures τ_d and τ_c to give weights to the delay and count cues:

$$P(n) = \frac{\exp(v_{dn}/\tau_d + v_{cn}/\tau_c)}{\sum_{n' \in N} \exp(v_{dn'}/\tau_d + v_{cn'}/\tau_c)} \tag{D.2}$$

E Model Performance

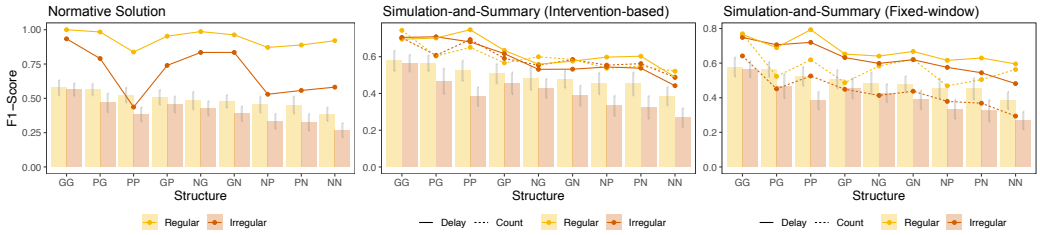

Figure E.1: Models' F1-score under different structures of experimental stimuli. Bars in the background indicate human performance.

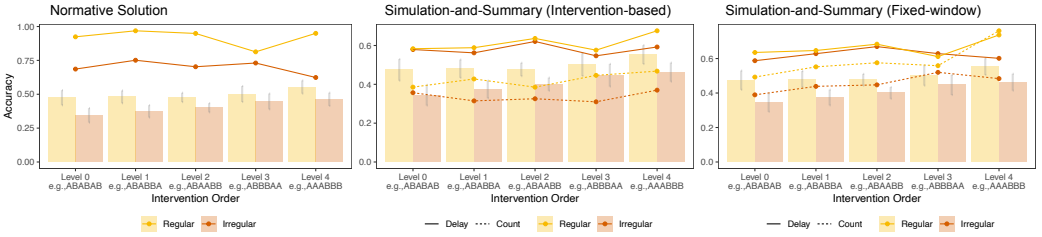

Figure E.2: Models' judgment accuracy under different intervention orders of experimental stimuli. Bars in the background indicate human performance.

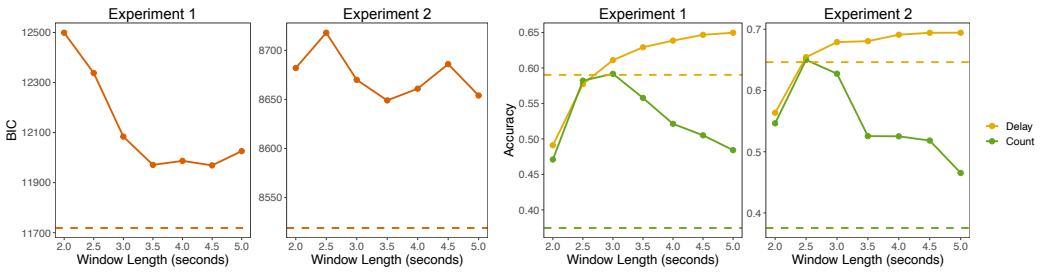

Figure E.3: BIC and model accuracy under different fixed window lengths of simulation-and-summary models. Horizontal dashed lines indicate cases of intervention-based segmentation.