## 1 Summary of supplementary experiments

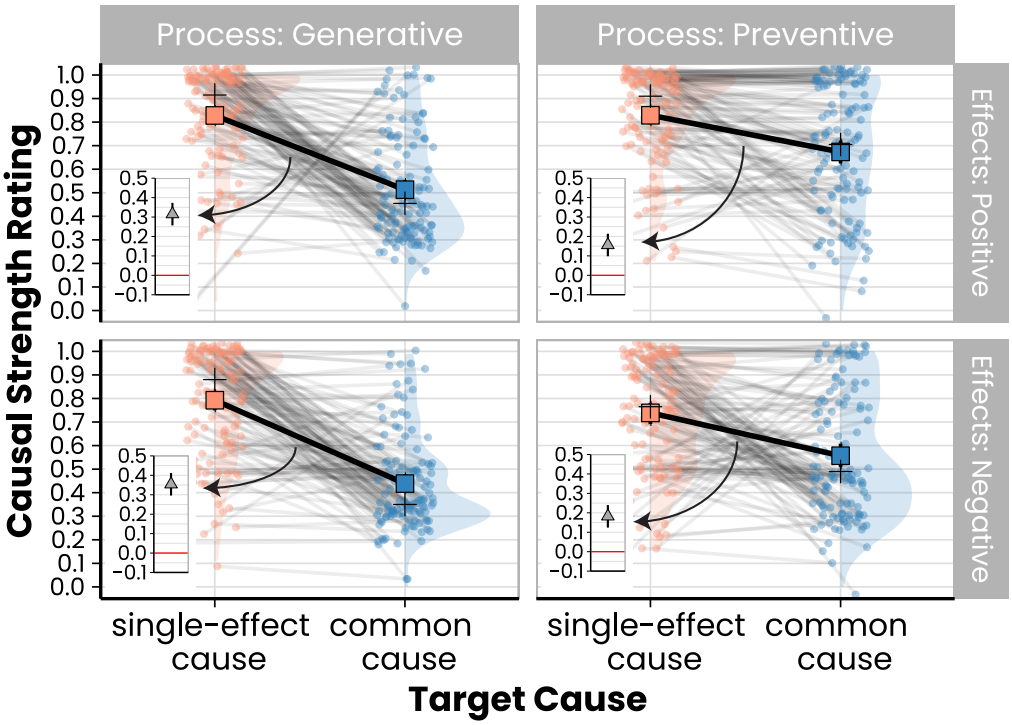
In a first supplementary study (Supplementary Study 1) we tested one prediction of Sussman and Oppenheimer's (2020) account. Sussman and Oppenheimer (2020) predicted that the influence of causal structure on causal strength intuitions depend on effect valence, that is, whether the effect(s) are perceived to be positive or negative. Our theory of perceived causal strength dilution does not predict an influence of effect valence. As we mentioned, a problem in Sussman and Oppenheimer's studies was that effect valence and type of causal process were confounded: the causes that led to positive effects were *preventive* causes that made existing negative states disappear (e.g., painful symptoms), whereas the causes that led to negative effects were *generative* causes that produced *previously absent* negative states (e.g., painful symptoms). This confounding of causality type and effect valence allowed subjects in the negative effects conditions, but not in the positive effects conditions, to infer that the (previously absent) negative effects generated by the common cause could be mutually amplifying each other. For example, subjects tended to reason that when a shaving cream causes ingrown hair, dry skin, and skin irritation, these different symptoms might also worsen *each other*. Also, in the test scenarios that were used, mutual amplification among the effects is a plausible assumption. Thus, when subjects rated the expected magnitude of change of a target symptom given the presence of either the common cause or the single-effect cause, their increased ratings for the common cause might have been driven largely by their inferring this mutual amplification amongst the effects of the common cause. We think that outcome valence per se does not influence perceived causal strength dilution.

Our experiment (testing $N = 480$ participants) had a 2 (causality type: generative vs. preventive; between subjects) $\times$ 2 (effect valence: positive vs. negative; between subjects) $\times$ 2 (causal structure: common cause vs. single-effect cause; within-subject) mixed design. Also, to avoid that subjects rely on background knowledge, we used our fictitious alien-crystal scenario. The effects this time were fictitious blood substances ("Sonin", "Pixin", and "Xantan"). We used these fictitious blood substances as effects because we assumed that subjects in this case might be less inclined to infer mutually amplifying links among them. The crystals that aliens eat were described to either produce (generative causality) or to decompose (preventive causality) the different fictitious blood substances. To manipulate effect valence, the blood substances were described either as healthy (positive effects) or unhealthy (negative effects). An example for the causal strength test questions subjects answered is: "According to your intuition, if an alien eats a blue [red] crystal, how strongly will this stimulate its body to produce [decompose] the blood substance Sonin [Pixin; Xantan]?" Ratings were provided on a continuous slider with the endpoints labeled "not at all" and "maximally".

Subjects' causal strength ratings in the different conditions are shown in Fig. 1. As can be seen there, we did not observe a reversal of the strength dilution effect: irrespective of whether the causes led to positive or negative effects, subjects tended to report causal strength dilution. A mixed ANOVA yielded a strong main effect of cause type (common cause vs. single effect cause), $F(1, 432) = 290.57$, $p < .001$, $\eta^2_{ges} = .240$, but, contrary to what Sussman and Oppenheimer's (2020) account predicts, the factor effect valence (positive vs. negative effects) did not significantly moderate the strength dilution effect. However, as

**Figure 1**

*Subjects' causal strength ratings in our Supplementary Study 1.*




*Note. Squares denote means and "+" denote medians. All error bars represent 95% CIs. Jittered dots and lines show subjects' individual ratings, and density plots show their distribution. The difference plot shows the estimated mean change.*

can be seen from the difference plots in Fig. 1, causal strength dilution tended to be slightly weaker when the causes were preventive instead of generative. This interaction between causal structure (common cause vs. single effect cause) and causality type (generative causality vs. preventive causality) was significant, $F(1, 432) = 31.46, p < .001, \eta^2_{ges} = .033$.

In a second supplementary study (Supplementary Study 2), we tested whether causal strength dilution also occurs when the causal strength test queries ask for the probability of the occurrence of an effect rather than for the magnitude of change. In our Experiment 1 we used probabilistic test queries in the conditions with binary variables. Our theory predicts an absence of strength dilution for binary variables, but an alternative explanation for this finding in Experiment 1 is that the effect does not occur with probabilistic test queries. Our second supplementary study addresses this problem by testing naturally continuous variables but using a probabilistic phrasing of the test queries. Furthermore, we also wanted to test whether perceived strength dilution still can be observed in situations in which reasoners only learn about one cause structure (common cause vs. single-effect cause) and thus cannot directly compare the two. We directly contrasted contexts in which subjects were jointly presented with both types of causes or with only one of them (joint vs. isolated presentation format).

The experiment (testing $N = 720$ subjects) had a 2 (presentation format: joint pre-

sentation of single-effect and common cause vs. presentation of only one of the two types of causes) × 2 (target cause: single-effect vs. common cause) between-subjects design. We used the alien-crystal scenario again. The experimental materials and procedure were largely identical to those in Experiment 2, except that half of the subjects (in the separate evaluation condition) learned about only one crystal that the aliens were described to eat. The probabilistic test questions read: "We now would like to get to know your intuition about the causal strength with which eating blue [red] crystals causes improved night vision [joint flexibility; alertness]. To express your intuition about causal strength, please answer the following question: What do you think is the probability with which eating red [blue] crystals leads to improved night vision [joint flexibility; alertness]?". Responses were given on an eleven-point rating scale with endpoints labeled "It never leads to improved night vision" and "It always leads to improved night vision".

We found that subjects again provided estimates indicating causal strength dilution with our probability strength query. Moreover, subjects tended show a dilution effect in both presentation format conditions, although the effect was slightly smaller when subjects learned about only one cause. In the condition in which subjects learned about both causes, the effect was $M_{diff} = M_{sec} - M_{cc} = 0.77 - 0.60 = 0.17$, 95% CI [0.12; 0.22]; $d = 0.71$, 95% CI [0.49; 0.92]; planned contrast: $t(716) = 6.99$, $p_{one-sided} < .001$. In the condition in which subjects learned about only one of the two types of causes the dilution effect was $M_{diff} = M_{sec} - M_{cc} = 0.67 - 0.60 = 0.07$, 95% CI [0.02; 0.11]; $d = 0.30$, 95% CI [0.09; 0.51]; planned contrast: $t(716) = 2.72$, $p_{one-sided} = .0034$. The difference of perceived strength dilution in the two condition was $\Delta_{diff} = 0.10$, 95% CI [0.04; 0.17]; planned contrast: $t(716) = 3.02$, $p_{one-sided} < .0013$. In sum, the findings of this study demonstrate that the occurrence of perceived causal strength dilution does not require a direct comparison of causes with different causal scopes, and it also shows that the effect can be obtained with probabilistic test queries.

Our previous studies compared one common cause with one single-effect cause. Another prediction that we tested in a third supplementary study (Supplementary Study 3; $N = 144$) is that perceived strength dilution should monotonically increase with the number of effects served by a common cause. We tested and confirmed this prediction using our alien crystal scenario with an additional third common cause of seven different effects. As predicted, causal strength ratings decreased with the number of effects of a cause, $M_{SEC} = 0.88$, 95% CI [0.849, 0.920], $M_{CC3} = 0.56$, 95% CI [0.52, 0.60], and $M_{CC7} = 0.41$, 95% CI [0.36, 0.46].

Two final supplementary studies (Supplementary Study 4a, $N = 216$, and 4b, $N = 216$) tested binary variables again. The reason why we conducted these studies was that Experiment 2 tested binary variables only in an abstract scenario, and we wanted to see whether the reduced tendency to perceive causal strength dilution in binary variable cases also extends to more specific scenarios. The scenario we used was one about binary light switches and binary LED lamps. In the single-effect case, a light switch was described to be connected to a single LED light. In the common-cause case, a light switch was described to be connected to three effects. We also added a control case in which a switch was not connected to its LED light (non-cause). Subjects learned that the light switch as well as the connected LED lights can only take on two possible states, "off" and "on". In the Supplementary Study 4b, subjects additionally learned that the light switches are

probabilistic. The switches were connected to tiny roulette wheels and the electrical signal was transmitted to the LED lights only with a 75% chance.

Corroborating our findings from Experiment 2, we only found very weak strength dilution effects in these binary scenarios. In Supplementary Study 4a, the effect was $M_{diff} = 0.954 - 0.948 = 0.0065$, 95% CI [−0.034, 0.047], $d = 0.033$, 95% CI [−0.155, 0.222]. Subjects' strength rating for the control case in which the switch box was not connected to any LED was $M_{nocause} = 0.04$, 95% CI [0.020, 0.0635], indicating that subjects correctly understood the instructed devices. We also found no pronounced strength dilution effect in Supplementary Study 4b. The difference between the ratings estimating the size of the dilution effect was $M_{diff} = 0.78 - 0.75 = 0.037$, 95% CI [−0.01, 0.08], $d = 0.17$, 95% CI [−0.02, 0.36].

The results of Supplementary Study 4a and 4b provide further evidence that causal strength dilution relies on a mental representation of the causal variables as being fundamentally continuous. When the causal variables are mentally represented as being genuinely binary (on-off or present-absent), reasoners do not seem to expect causal strength dilution.

## 2 References

Sussman, A. B., & Oppenheimer, D. M. (2020). The effect of effects on effectiveness: A boon-bane asymmetry. *Cognition*, *199*, 104240.