# Evaluating General versus Singular Causal Prevention

**Simon Stephan**[1] (simon.stephan@psych.uni-goettingen.de), **Sarah Placì**[2] (sarah.placi@unitn.it)
**Michael R. Waldmann**[1] (michael.waldmann@bio.uni-goettingen.de)
[1]Department of Psychology, University of Göttingen, Germany
[2]Animal Brain Cognition Group, University of Trento, Italy

## Abstract

Most psychological studies focused on how people reason about generative causation, in which a cause produces an effect. We here study the prevention of effects both on the general and singular level. A general prevention query might ask how strongly a vaccine is expected to reduce the risk of contracting COVID-19, whereas a singular prevention query might ask whether the absence of COVID-19 in a specific vaccinated person actually resulted from this person's vaccination. We propose a computational model answering how knowledge about the general strength of a preventive cause can be used to assess whether a preventive link is instantiated in a singular case. We also discuss how psychological models of causal strength learning relate to mathematical models of vaccination efficacy used in medical research. The results of an experiment suggest that many, but not all people differentiate between preventive strength and singular prevention queries, in line with the formal model.

**Keywords:** prevention; causal strength; vaccination efficacy; general causation; singular causation; computational modeling

On December 31, 2019, the World Health Organization's (WHO) country office in China picked up a media statement by the Wuhan Municipal Health Commission on cases of viral pneumonia in Wuhan. After only a short time, it became apparent that these cases of illness were caused by a novel virus, *SARS-CoV-2*. The virus turned out to be spreading so rapidly that the outbreak had to be declared a pandemic on March 11, 2020 (see `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline?`).

One of the most effective means to combat viruses, such as SARS-CoV-2, is vaccination. Accordingly, great efforts have been made to develop vaccines against SARS-CoV-2. The first effective vaccines were developed and licensed within just one year. A key parameter in the development of vaccines is their *preventive efficacy* (see, e.g., Orenstein et al., 1985), which typically is understood as a vaccine's capacity to reduce the risk of severe disease progression, assessed in the context of a clinical trial. For example, the vaccines of Moderna and BionTech-Pfizer, which have already been approved by several countries, are expected to have a preventive efficacy of about 95 percent (see, e.g., Zimmer, 2020).

From a psychological perspective, the urgent search for vaccines during a pandemic forcefully illustrates that it is important to understand how people learn and think about *preventive causal relations*. In preventive causal relations, the occurrence of a target cause (e.g., getting vaccinated) is associated with the *absence* of an effect (e.g., not contracting a disease). Although preventive causation is captured by different theoretical frameworks of causal induction (e.g., Cheng, 1997; Goldvarg & Johnson-Laird, 2001; Wolff, 2007), psychological studies on causal learning and reasoning mostly focused on how people learn and think about *generative causal relations* (e.g., Griffiths & Tenenbaum, 2005) (see also Waldmann, 2017, for overviews), in which the *occurrence* of a target cause is associated with the *occurrence* of a target effect (but see, e.g., Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008; Walsh & Sloman, 2011; Wolff, 2007, for exceptions).

We here investigate how reasoners acquire and apply their knowledge about *preventive causal relations*. In particular, we focus on the question of how knowledge about the general *strength* of a preventive causal relation can be used to determine the probability with which this preventive relation is *actually* instantiated in a *singular case*. Knowledge about the general strength of preventive causal relations is, among other things, important to plan successful interventions (e.g., should vaccine A or B be selected for the vaccination campaign?). Being able to assess how likely it is that a preventive measure has actually worked in a singular case is also important, however. For example, people who believe that vaccination actually prevented them from contracting a disease may be more willing to get vaccinated against another disease in the future. In contrast, vaccinated people who believe that it was not the vaccination (but something else) that actually prevented them from contracting the disease might be less willing to do so.

To answer the question of how the presence of a singular instance of prevention can be assessed based on knowledge about the general strength of a preventive cause, we propose a novel equation inspired by the generalized power PC model of singular causation judgments (Stephan, Mayrhofer, & Waldmann, 2020; Stephan & Waldmann, 2018). The model has previously been used to model singular causation judgments in generative cases, but not singular prevention judgments. Staying close to the introductory scenario, our running example will be the vaccination against a novel disease. Since a crucial component of our model of singular prevention is knowledge about the general strength of a preventive cause, a further theoretical goal is to compare and connect mathematical models used in medicine to assess vaccination efficacy with psychological computational models of causal strength induction. We also present the results of a first experiment, in which we asked subjects to answer either general preventive efficacy or singular prevention queries. To foreshadow our results, in line with the formal analysis, subjects tended to give different answers to the two types of queries. The general and singular prevention judgments of many subjects were predicted well by the formal models.

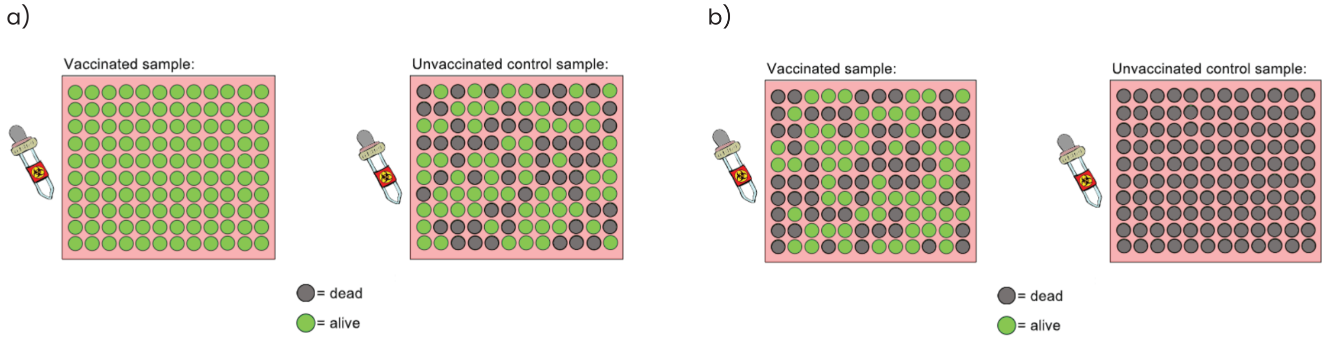

Figure 1: Two example contingency data sets. In a) vaccination is a sufficient but not a necessary preventer of the disease. In b) vaccination is a necessary but not a sufficient preventer.

However, we also found that some subjects appear to have confused singular prevention queries with general preventive strength queries. We discuss theoretical and empirical implications of our findings and conclude with an outlook on future studies.

## Estimating a vaccine's preventive efficacy

The standard mathematical formula used by medical researchers to determine the preventive efficacy of a vaccine has been developed by Yule and Greenwood (1915), who evaluated the outcomes of numerous clinical studies investigating the success of vaccination (called *preventive inoculation* at the time) against typhoid and cholera. The version of their mathematical formula most often used today was proposed by Orenstein et al. (1985). It is given by the following equation:

$$VE = \frac{ARU - ARV}{ARU} \cdot 100. \tag{1}$$

$ARU$ denotes the attack rate among unvaccinated individuals, which is the proportion of cases of the disease that occurred during a defined period of time in the population of unvaccinated individuals. $ARV$ denotes the attack rate among vaccinated individuals. It represents the proportion of cases of the disease that during the same period occurred in the vaccinated population. To express $VE$ as a percentage, the fraction is multiplied by 100.

Fig. 1 shows two graphical illustrations that are useful to demonstrate how Equation 1 works. The left panel in each figure represents a population of 120 vaccinated individuals and the right panel represents a population of 120 unvaccinated individuals. The effect in this example is "contraction of a fatal disease caused by a bacillus". Individuals who survived and who died from the disease are depicted as green and grey circles, respectively. The pipettes with the warning label next to each panel illustrate that both the vaccinated as well as the unvaccinated populations had been exposed to the bacillus. In Fig. 1a, the attack rate among unvaccinated individuals ($ARU$) is $60/120 = 0.5$, whereas the attack rate among vaccinated individuals ($ARV$) is $0/120 = 0$. By contrast, in Fig. 1b all unvaccinated individuals have died, $ARU = 120/120 = 1$, while the attack rate among vaccinated individuals ($ARV$) is $60/120 = 0.5$. The numerator of Equation 1 measures how much the probability of the effect decreases in the presence compared to the absence of vaccination. The attack rate reduction given vaccination is identical in Fig. 1a and b, $ARU - ARV = 0.5$, or 50%. To control for the fact that some individuals may survive even without vaccination, the observed attack rate reduction is then restricted to the proportion of individuals in whom the vaccination can be expected to make a difference. As $ARU$ is an estimator of this proportion of individuals, this is achieved by dividing the attack rate reduction through $ARU$. As a result, despite identical reduction rates, the estimated vaccination efficacy differs between Fig.1a and b. In Fig. 1a, it is $VE = \frac{ARU-ARV}{ARU} \cdot 100 = \frac{0.5-0}{0.5} \cdot 100 = 100\%$. In Fig.1b, $VE$ is equal to the reduction rate, $VE = \frac{ARU-ARV}{ARU} \cdot 100 = \frac{1.0-0.05}{1.0} \cdot 100 = 50\%$.

The examples illustrate that Equation 1 measures the degree of *sufficiency* of a preventive cause. In Fig. 1a, the vaccine appears to be perfectly effective, captured by $VE = 100\%$, whereas this does not seem to be the case for the vaccine in Fig. 1b, despite identical observed reduction rates.

It has already been established by Yule and Greenwood (1915) that Equation 1 is a reliable estimator for a vaccine's preventive efficacy only if certain conditions are met (Yule & Greenwood, 1915, pp. 115):

1. The persons must be, *in all material respects*, alike.

Under *all material respects* Yule and Greenwood (1915) understand factors that might affect the liability to the disease or, if the outcome is the fatality rate as in our example, the probability to die from the disease. These factors may include age, medical condition, or the history of prior infections.

2. The effective exposure to the disease must be identical in the case of inoculated and unvaccinated people.

3. Inoculation and disease should have occurred independently.

The second criterion would be violated, for example, if vaccinated individuals were systematically exposed to a less potent variant of the virus than unvaccinated individuals. In this case, Equation 1 can be expected to overestimate $VE$.

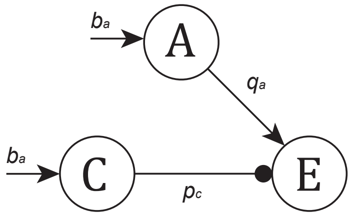

Figure 2: Causal Bayes net in which $C$ and $A$ combine according to a noisy-AND-NOT function. $C$ represents a preventive cause of effect $E$. $A$ summarizes all generative causes of $E$. $b_c$ and $b_a$ denote the causes' base rates. $p_c$ and $q_a$ denote the causes' preventive and generative strength, respectively.

The third criterion would be violated, for example, if the sample of unvaccinated subjects contained a higher number of individuals who already had contracted (and survived) the disease. As the attack rate among unvaccinated patients ($ARU$) can be expected to be reduced in this case, Equation 1 would underestimate $VE$.

## A causal Bayes net representation of vaccination efficacy

From a causal Bayes net perspective (Griffiths & Tenenbaum, 2005; Pearl, 2000), the assumptions under which Equation 1 yields a reliable estimate of a vaccine's general preventive efficacy $VE$ instantiate a common-effect network in which the causes of the target effect combine according to a noisy-AND-NOT function. An illustration is shown in Fig. 2. $C$ (e.g., a vaccine) represents a preventive cause of $E$ (e.g., a disease), and $A$ represents a conglomerate of generative causes of the effect (or only a single generative cause, such as a novel virus). Under a noisy-AND-NOT parameterization, $C$ prevents $E$ with strength $p_c$ and $A$ generates $E$ with strength $q_a$ (or $b_a \cdot q_a$ if $A$ is unobserved). It is also assumed that $C$ and $A$, when simultaneously present, influence $E$ independently, that a cause's strength does not depend on its base rate, and that $E$ does not occur unless it is caused (Cheng, 1997). Under these assumptions, the generating function of the effect is: $P(e^+|a,c;q_a,p_c) = b_a q_a - b_a q_a b_c p_c$. As Cheng (1997) showed, the preventive strength (or power) $p_c$ of $C$ can then be estimated based on the observed contingency between $C$ and $E$ by:

$$p_c = -\frac{P(e^+|c^+) - P(e^+|c^-)}{P(e^+|c^-)} = -\frac{\Delta P}{P(e^+|c^-)}. \qquad (2)$$

Equations 1 and 2 can be directly related. If $E$ represents the contraction of a fatal disease caused by a virus and $C$ represents vaccination, then $P(e^+|c^+)$ and $P(e^+|c^-)$ correspond to $ARV$ and $ARU$, respectively. Thus, $VE = p_c \cdot 100$ or $p_c = \frac{VE}{100}$ in this case. Furthermore, both equations express that a candidate cause's preventive strength can only be determined if the probability of the effect in the absence of the candidate cause ($P(e^+|c^-)$ or $ARU$) is $> 0$; both equations are undefined for $ARU = P(e^+|c^-) = 0$. One reason why the efficacy of Corona vaccines could be determined relatively fast is that (sadly) $ARU$ was notably larger than 0 in the beginning of the pandemic.

## Estimating the probability of actual prevention

We now show how, under the above conditions, $p_c$ (or $VE$) can be used to estimate the probability with which a preventive cause has *actually* prevented the effect in a singular case. To accomplish this, we build on the generalized power PC model of singular causation judgments (Stephan et al., 2020; Stephan & Waldmann, 2018), which models singular causation judgments in generative causal scenarios. In the generative case, in which $C$ is assumed to produce $E$ with strength $q_c$, the model provides answers to queries like "How likely is it that $C$ caused $E$ in this particular case in which $c^+$ and $e^+$ actually co-occurred?" by estimating $P(c^+ \rightarrow e^+|c^+,e^+)$, which is given by:

$$P(c^+ \rightarrow e^+|c^+,e^+) = \frac{q_c - q_c b_a q_a \alpha}{q_c + b_a q_a - q_c b_a q_a} = \frac{q_c - q_c b_a q_a \alpha}{P(e^+|c^+)}. \qquad (3)$$

The numerator term $q_c b_a q_a \alpha$ represents the probability with which the target cause $c^+$ was *causally preempted* by a competing generative cause $a^+$ on an occasion, which needs to be subtracted from the target cause's strength $q_c$. Since causal preemption can occur only on occasions on which competing generative causes are actually strong enough to produce $e^+$, one part of this term is $q_c b_a q_a$, which identifies these occasions. On these occasions it needs to be specified whether $a^+$ causally preempted $c^+$. The parameter $\alpha$ determines the proportion of cases on which this is the case. According to the theory, $\alpha$ can be determined based on temporal information (see Stephan et al., 2020).

In a preventive context like the vaccination scenario, we may accordingly ask "What is the probability that the vaccination of this individual actually prevented this individual from contracting the fatal disease caused by the virus?". Importantly, unlike previous studies that solely focused on general preventive causal strength (e.g., Lu et al., 2008), we here distinguish between cases in which a cause prevents an effect from happening, i.e., cases in which an effect is initially absent and remains absent, and cases in which a preventive cause makes an effect disappear. The notation we introduce for cases in which a cause prevents an effect from happening, which are the ones we focus on, is $P(c^+ \multimap e^+|c^+,e^-)$. By contrast, the notation we suggest for cases in which the presence of a preventive cause initiates a change of the effect's status from present to absent is $P(c^+ \rightarrow e^-|c^+,e^-)$. A key difference between these two cases is that a singular instantiation of prevention in which the target cause keeps the effect from happening ($c^+ \multimap e^+$) requires that (1) the generative cause of the effect is present and (2) sufficiently strong to produce the effect. The idea here is that actual prevention cannot occur in a specific case unless there actually is a generative cause of the effect present that would lead to the occurrence of effect ($e^+$) if the preventer was not present. By contrast, in cases in which a preventive cause leads to a change in the effect's status from present to absent ($c^+ \rightarrow e^-$), the base rate and strength of the generative cause(s) that generated the effect ($e^+$) can be neglected for the computation of the probability of actual prevention of $e$ by $c$. Written in terms of the parameters of the underlying causal Bayes net, the equation estimating $P(c^+ \multimap e^+|c^+, e^-)$ is:

$$P(c^+ \multimap e^+|c^+, e^-) \;=\; \frac{p_c b_a q_a}{1-(b_a q_a - p_c b_a q_a)}. \tag{4}$$

The product $p_c b_a q_a$ in the numerator of Equation 4 represents the relative frequency of cases in which $C$ is strong enough to prevent $E$ if the generative cause $A$ is simultaneously strong enough to generate $E$. This term is the theoretically most important element of the equation. It expresses the assumption that a preventive target cause can actually have prevented the target effect only on occasions on which a generative cause of that effect is present and strong enough to generate the effect. Expressed in counterfactual terminology, a preventive cause actually prevented an effect only if a present generative cause would have produced the effect if the preventive cause had been absent. Furthermore, since according to Fig. 2 the target cause $C$ is not competing with further explicit preventive causes, the problem of causal preemption does not occur. The influence of additional unobserved preventive causes is assumed to be expressed implicitly in $A$'s parameters (and thus $ARU$). For example, if $ARU\; = 1$, one can conclude that no preventive causes other than $C$ exist. The denominator of Equation 4 represents the relative frequency of cases in which the target effect actually remained absent in the target cause's presence. Equation 4 can also be expressed purely in terms of observable probabilities:

$$P(c^+ \multimap e^+|c^+, e^-) \;=\; \frac{P(e^-|c^+) - P(e^-|c^-)}{P(e^-|c^+)}, \tag{5}$$

or in a form using the medical terminology of Equation 1:

$$P(v^+ \multimap a^+|v^+, a^-) = \frac{VE \cdot ARU}{1-ARV}. \tag{6}$$

In the latter equation, $v$ denotes *vaccination* and $a$ denotes *attack* by the disease. To illustrate these equations, we return to Fig.1. We consider a randomly sampled vaccinated individual who survived. In Fig.1a, the probability that the vaccination actually prevented a fatal attack in this individual is $P(v^+ \multimap a^+|v^+, a^-) \;=\; \frac{1 \cdot 0.5}{1-0} \;= 0.5$. Equation 6 takes into account that 50% of the individuals are not attacked by the disease even without vaccination (e.g., due to natural immunity). It predicts that vaccination did not actually prevent the disease in such a case, even though it is generally a hundred percent effective. Thus, in this case vaccination is not necessary to stay healthy. A different prediction is obtained for the example in Fig.1b. In this case $P(v^+ \multimap a^+|v^+, a^-) \;=\; \frac{0.5 \cdot 1}{1-0.5} \;= 1.0$. Since all unvaccinated individuals are attacked by the disease, vaccination is necessary to stay alive. The model predicts that every healthy vaccinated individual must therefore have been protected by the vaccination, even though vaccination does not guarantee health.

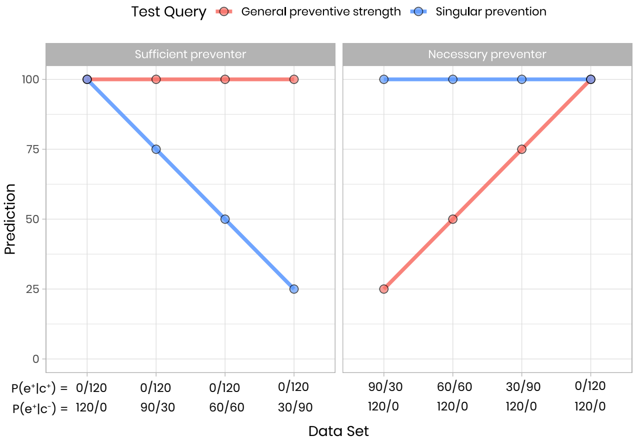

Figure 3: Predictions for different data sets tested in the experiment. Red lines show preventive strength values (Equations 1 and 2). Blue lines show singular prevention predictions (Equations 5 and 6).

Fig. 3 shows the predictions for different data sets. Red lines show the general preventive strength as computed by Equations 1 and 2. Blue lines show the predictions for actual/ singular prevention as computed by Equations 4, 5, and 6. The left panel displays data sets in which the target cause is a sufficient but not a necessary preventer of the effect, whereas the right panel displays data sets in which the cause is a necessary but not a sufficient preventer.

## Experiment

The goal of the experiment was to evaluate to which extent lay people answer general vs. actual preventive strength queries as predicted by the models. The experimental data and all materials can be accessed at `https://simonstephan31.github.io/singular_prevention_proceedings`. The different contingency data sets we tested are those shown in Fig. 3, which we used because they clearly dissociate between the two different types of queries. The experimental scenario was about a group of biologists conducting laboratory studies with mice to test the efficacy of a candidate vaccine against different strains of a novel dangerous bacillus. The scenario described a randomized controlled trial (RCT) in which different random samples of mice either were or were not vaccinated against a certain strain of the novel bacillus. Subjects learned that they would see the results of four experiments conducted by the biologists in which they tested the candidate vaccine against four different strains of the bacillus. Subjects were told that the results of each experiment would be presented one after another in the form of short animations. The illustrations in Fig. 1 show snapshots of the animations. Finally, subjects were informed about the type of test query (a general preventive vs. a singular prevention query) that they would be asked for each of the observed data sets.

### Methods

**Participants** One hundred and four subjects ($M_{age} \;= 37.34$ years, $Range_{age} \;= 18-77$ years, 56 female, 47 male, 1 non-binary) recruited via Prolific (`www.prolific.co`) par-

Table 1: Experimental results

| | Sufficient Preventer | | | | Necessary Preventer | | | |
|---|---|---|---|---|---|---|---|---|
| | 0/120<br>120/0 | 0/120<br>90/30 | 0/120<br>60/60 | 0/120<br>30/90 | 90/30<br>120/0 | 60/60<br>120/0 | 30/90<br>120/0 | 0/120<br>120/0 |
| $M_{strength}$ | 97.9 | 93 | 88.3 | 87.2 | 28.1 | 54.2 | 70.5 | 99.6 |
| *SD* | 6.5 | 15.9 | 20.1 | 26.5 | 12.1 | 7.2 | 10.1 | 1.1 |
| 95% CI | [90.9; 100] | [86.0; 100] | [81.3; 95.3] | [80.2; 94.2] | [21.1; 35.1] | [47.1; 61.2] | [63.5; 77.5] | [92.6; 100] |
| $M_{singular}$ | 97.5 | 89.9 | 81.9 | 66.8 | 63.2 | 76.8 | 88.8 | 94.8 |
| *SD* | 8.6 | 11.0 | 18.1 | 30.6 | 32.6 | 22.4 | 15.0 | 18.0 |
| 95% CI | [90.5;100] | [82.9; 96.9] | [74.9; 88.9] | [59.8; 73.8] | [56.2; 70.2] | [69.8; 83.8] | [81.8; 95.5] | [87.8; 100] |

ticipated in this online study and provided valid data. The inclusion criteria were a minimum age of 18 years, English as native language, and an approval rate obtained in previous studies of 90 percent. Subjects were asked not to participate via smartphone or tablet.

**Design, Materials, and Procedure** The study had a 2 (type of test query: general preventive strength vs. singular prevention; varied between subjects) × 2 (type of preventive cause: sufficient vs. necessary; varied between subjects) × 4 (contingency data set: four per condition as shown in Fig.3; varied within subject in random order).

A demo video of the study can be found at `https://tinyurl.com/32dbmtbt`. Subjects were first given some general information about the experiment. They then read the scenario description and received additional procedural information. During the learning and test phase they were shown four animations conveying the contingencies displayed in Fig. 3. The animations were presented on separate screens. Subjects saw that all mice were alive in the beginning. After two seconds, all mice were simultaneously exposed to the bacillus, which was indicated by pipettes that appeared and released the bacillus into the mice panels. An example animation can be viewed at `https://tinyurl.com/9cwvv38m`. Every animation lasted 16 seconds and subjects could watch them as often as they wanted.

The test question was shown below each animation. Subjects in the "general preventive strength query" condition were asked: "How effectively does the vaccine prevent mice from dying from the disease that can be caused by the investigated strain of bacteria? To rate the vaccine's effectivity, imagine a new group of 100 unvaccinated mice who all died from the disease caused by the studied strain of bacteria. Based on what you have learned, if these 100 mice had been vaccinated, how many do you think would have survived?" The wording followed formulations used in previous studies on causal strength learning (see, e.g., Lu et al., 2008). Ratings were provided on a continuous slider with endpoints labeled "None of them (0)" and "All of them (100)". Participants in the "singular prevention query" condition were asked: "Imagine one of the living mice is randomly selected from the vaccination group. Based on what you have learned, how confident are you that it actually was the vaccination that prevented this mouse from dying from the disease that can be caused by the studied strain of bacteria?" Ratings were provided on a continuous slider with endpoints labeled "Certain that it was not the vaccination that prevented the mouse from dying (0)" and "Certain that it was the vaccination that prevented the mouse from dying (100)". To encourage subjects to think thoroughly, we also asked them to write short explanations for their ratings.

## Results and Discussion

Table 1 and Fig. 4 show the results. The bold colored lines show the mean ratings (error bars denote 95% bootstrapped CIs) and the Jittered pale lines show subjects' individual ratings. Subjects tended to answer general preventive strength and singular prevention queries differently. The overall pattern are generally consistent with the predictions shown in Fig. 3. Subjects' preventive strength ratings largely followed Equations 1 and 2, whereas their singular prevention judgments were better explained by Equations 5 and 6. In line with the predictions, a mixed ANOVA (with Greenhouse-Geisser correction applied) yielded significant interaction effects between (1) "Type of test query" and "type of preventive cause", $F(1,100) = 24.10$, $p < .001$, $\eta_p^2 = .194$, and (2) between "Type of test query" and "contingency data set", $F(2.03, 202.63) = 20.22$, $p < .001$, $\eta_p^2 = .168$. The three-way interaction between all factors did not reach significance, however, $p = .127$.

Table 1 and Fig. 4 also show that we observed high interindividual variability in subjects' ratings, particularly in

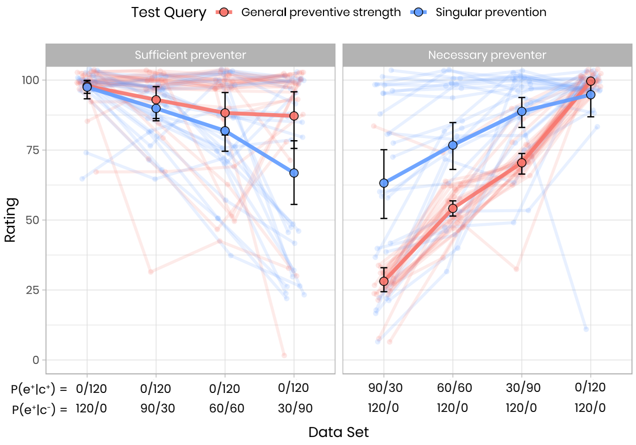

Figure 4: Experimental results. Bold lines show the mean ratings (error bars show 95% bootstrapped CIs). The jittered thin pale lines show subjects' individual ratings.

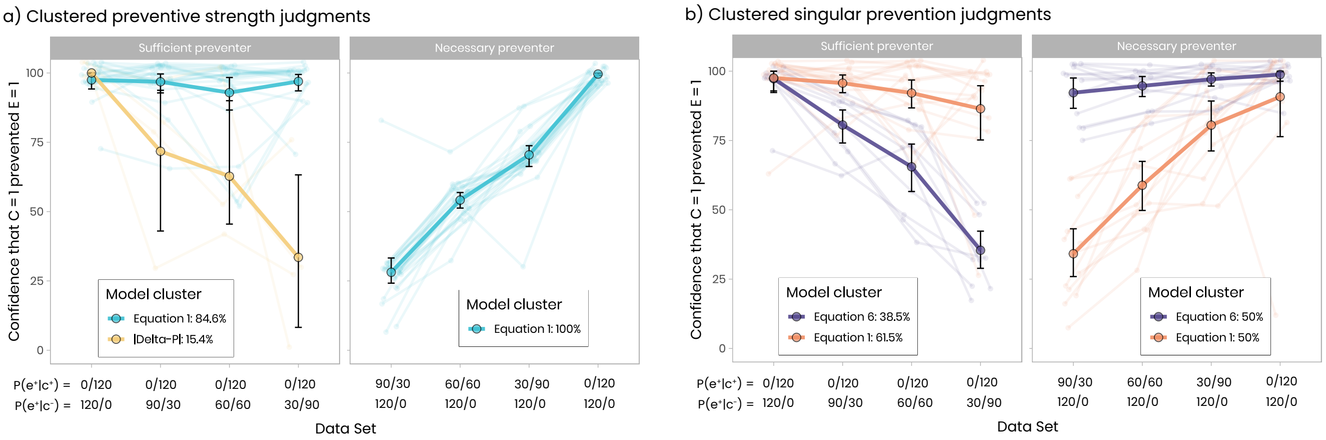

Figure 5: Results of the model cluster analysis. Bold lines show mean ratings of the clusters (error bars denote 95% bootstr. CIs).

their singular prevention ratings (blue lines). The pattern indicates two distinct subgroups. While the singular prevention ratings of one subgroup closely followed the predictions of Equations 4, 5, and 6, a second group tended to answer the singular prevention queries as if they were responding to the general preventive strength query. As for the generative prevention judgments, the red lines in the "sufficient preventer" condition indicate that a small subgroup of subjects tended to report the observed probabilistic difference ($ARU - ARV$ or $|\Delta P|$) instead of $VE$.

To further examine the individual rating patterns, we conducted a model-based cluster analysis. In the "general preventive strength query" condition, the included models were Equation 1 and $ARU - ARV$ (or $|\Delta P|$). In the "singular prevention query" condition, we included Equations 6 and 1. Subjects were assigned to model clusters based on the minimum mean distance of their ratings from the predictions of the models. Fig. 5 shows that the general preventive strength ratings (a) of most subjects were best predicted by $VE$ (or $p_c$). The ratings of a small group followed $ARU - ARV$ (or $|\Delta P|$). As for the "singular prevention query" condition, the results corroborated the rating pattern indicated in Fig. 4. In the condition in which subjects saw that vaccination was a necessary preventer, half of the subjects gave singular prevention ratings in line with Equation 6. The other half seems to have responded with the general preventive strength ($VE$ or $p_c$) of the cause. In the "sufficient preventer" condition, the proportions were 40 and 60 percent, respectively. A review of subjects' explanations corroborated the cluster analysis. Prototypical explanations given by subjects assigned to the "general preventive strength" cluster were: (1) "The vaccination didn't help much because there was still a lot of deaths" or (2) "From the videos it looks as if the mice have a 50% chance of survival if vaccinated". The explanations of subjects in the "Equation 6" cluster tended to express the logic behind the equation. Two examples are: (1) "100% fatalities in the control group, but something helped 25% of the vaccinated mice to survive. Not a tremendous success rate but nevertheless, in the absence of other known factors the vaccine seems likely to be the reason", and (2) "All of the mice that were not vaccinated died so I concluded that any mouse that survived in the vaccinated group was a result of being vaccinated". Some of the explanations suggest that some subjects also doubted the general efficacy of the vaccine, especially for the data sets in which its preventive strength was low. These subjects seemed to have had a strong prior for high preventive strength, which does not seem implausible in a vaccination scenario. As Stephan and Waldmann (2018) have shown, high confidence in the existence of a general causal link is a prerequisite for high confidence in a singular instantiation of this link. In sum, the results suggest that many reasoners differentiate between general preventive strength and singular prevention queries, but responses to the latter type of query were more heterogeneous than those to the first.

## General Discussion

In the present research we provided a formal answer to the question of how singular instances of causal prevention can be assessed in light of knowledge about the general strength of a preventive causal relation. The new model we presented computes the probability with which a preventive target cause $C$ prevented a target effect $E$ from happening given that $C$ occurred and $E$ failed to occur, $P(c^+ \multimap e^+ | c^+, e^-)$. We also showed that the mathematical formula used by medical researchers to determine vaccination efficacy can directly be related to computational models of causal strength learning developed by psychologists and computer scientists. We furthermore showed how our equation of actual singular prevention (Equation 4), expressed in the form of causal Bayes net parameters, can be translated into the terminology used by medical researchers (Equation 6). Finally, we reported the results of a first empirical test of our model of actual prevention judgments.

The key assumption of our equation of singular prevention is that actual prevention can only take place on occasions on which a sufficiently strong preventive cause meets a sufficiently strong generative cause. Our experiment showed that lay people seem to differentiate between general preventive strength and singular prevention queries, roughly in accordance with the formal models. However, these initial results also suggest that singular prevention queries might be more difficult to answer than general preventive strength queries.

In future studies, we plan to present additional contingency data sets and to vary the base rate of the generative cause (e.g., the prevalence of the bacillus). This will allow us to ask subjects about test cases in which the preventive cause is present but the generative cause is absent. Our model predicts that subjects should rule out a singular instance of prevention in such cases. To generalize beyond vaccination scenarios, we also plan to test additional cases of prevention. Finally, we plan to test and compare scenarios in which a preventer keeps an effect from happening ($c^+ \multimap e^+$) with those in which a preventive cause makes an effect disappear ($c^+ \rightarrow e^-$).

**Acknowledgments** This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Grant WA 621/24-1.

The experimental materials and data can be accessed at `https://simonstephan31.github.io/singular_prevention_proceedings`.

## References

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*(2), 367–405.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*(4), 565–610.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*(4), 334–384.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*(4), 955–982.

Orenstein, W. A., Bernier, R. H., Dondero, T. J., Hinman, A. R., Marks, J. S., Bart, K. J., & Sirotkin, B. (1985). Field evaluation of vaccine efficacy. *Bulletin of the World Health Organization*, *63*(6), 1055–1068.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, England: Cambridge University Press.

Stephan, S., Mayrhofer, R., & Waldmann, M. R. (2020). Time and singular causation – a computational model. *Cognitive Science*, *44*(7), e12871.

Stephan, S., & Waldmann, M. R. (2018). Preemption in singular causation judgments: A computational model. *Topics in Cognitive Science*, *10*(1), 242–257.

Waldmann, M. R. (Ed.). (2017). *The Oxford handbook of causal reasoning*. New York: Oxford University Press.

Walsh, C. R., & Sloman, S. A. (2011). The meaning of cause and prevent: The role of causal mechanism. *Mind & Language*, *26*(1), 21–52.

Wolff, P. (2007). Representing causation. *Journal of experimental psychology: General*, *136*(1), 82–111.

Yule, G. U., & Greenwood, M. (1915). The statistics of anti-typhoid and anticholera innoculations, and the interpretation of such statistics in general. *Royal Society of Medicine Proceedings, Section of Epidemiology and State Medicine*, *8*, 113–194.

Zimmer, C. (2020, December 4). 2 companies say their vaccines are 95% effective. what does that mean? *The New York Times*. Retrieved 2021-04-16, from `https://www.nytimes.com/2020/11/20/health/covid-vaccine-95-effective.html`