Predictions of the generalized power PC model of singular causation judgments for Exp. 1

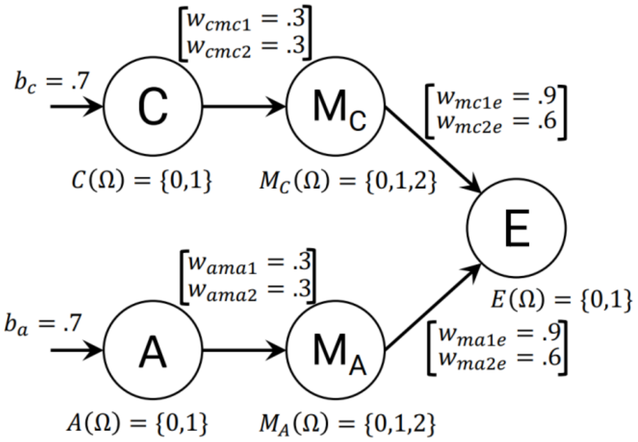


Figure 1. The underlying general causal model and its strength parameters

Fig. 1 shows the general causal model and the strength parameter. Nodes C, A, and E represent binary variables that can either be absent or present (0 vs. 1). In our experimental scenario C = 1 and A = 1 mean that the castles sent out a carrier pigeon. E = 1 means that an alarm occurred in the King's palace. M_C and M_A represent ternary variables. In our experimental scenarios, these nodes are the intermediate stations. A value of 0 means that the intermediate stations are inactive. A value of 1 means that an intermediate station is active and sends a telegraph. A value of 2 means that an intermediate station is active and sends a pony rider. The causal strength parameters w_{cmc1} and w_{ama1} denote the strengths with which C and A cause $M_C = 1$ and $M_A = 1$, respectively. The causal strength parameters w_{cmc2} and w_{ama2} denote the strengths with which C and A cause $M_C = 2$ and $M_A = 2$, respectively. The causal strength parameters w_{mc1e} and w_{ma1e} denote the strengths with which $M_C = 1$ and $M_A = 1$ cause E = 1. The causal strength parameters w_{mc2e} and w_{ma2e} denote the strengths with which $M_C = 2$ and $M_A = 2$ cause E = 1.

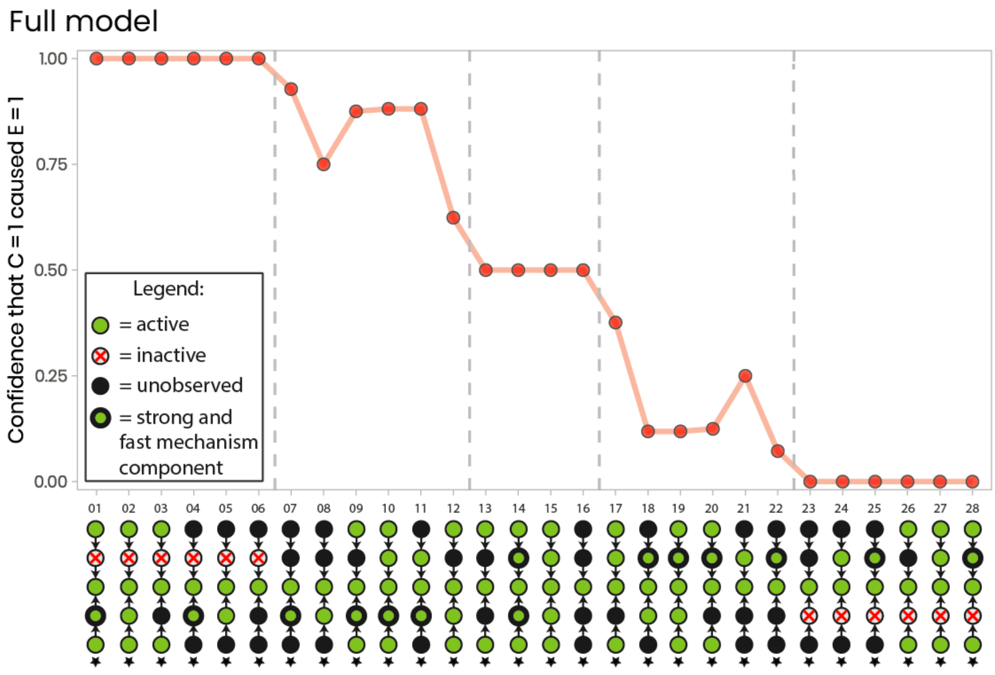


Figure 2. Test cases (x-axis) and model predictions.

Fig. 2 shows the model predictions for the 28 test cases of Experiment 1. The test cases are shown on the x-axis in the form of neuron diagrams. Green nodes with a regular contour represent active variables (state = 1). Black nodes represent unknown variable states (state = ?). Crossed nodes represent inactive variables (state = 0). Green nodes with a bold contour represent state = 2 of the intermediate variables. Nodes marked with a star represent the target cause.

To determine the probability that target cause c caused e, $P(c \rightarrow e|e,\ and\ othergivens)$, we need to determine the probability that c caused its mechanism variable, $P(c \rightarrow m_c|givens)$, and the probability that the mechanism variable cause the effect, $P(m_c \rightarrow e|e, and\ other\ givens)$. The probability that c caused e corresponds to the product of the probability that c caused its mechanism variable and the probability that the mechanism variable caused the effect.

Model predictions

<u>Segment 1: test cases 01 – 06</u>

As can be seen in Fig. 2, in all these test cases the intermediate variable of the alternative competing cause is absent (crossed nodes for M_A). Referring to our experimental scenario, the competing castle failed to activate its intermediate stations in these cases. This means that the competing castle's causal strength is 0. As result, our model predicts that the effect must have been caused by the target castle. We illustrate the model predictions using test case 01.

Givens: in test case 01, we observe that both castles send a pigeon, that an alarm occurred in the palace, that the competing castle's intermediate station is inactive, and that the intermediate station of the target castle sent a telegraph.

Step 1: The probability that C = 1 caused M_C = 1.

$$P(c \rightarrow m_{c=1}|c, m_{c=1}) = \frac{w_{cmc1} - w_{cmc1} b_a w_{amc1} \alpha}{w_{cmc1} + b_a w_{amc1} - w_{cmc1} b_a w_{amc1}}$$

Since M_C has no alternative cause, $b_a w_{amc1} \alpha = 0$. The equation thus reduces to 1.0:

$$P(c \rightarrow m_{c=1}|c, m_{c=1}) = \frac{w_{cmc1}}{w_{cmc1}} = 1.0$$

Step 2: The probability that M_C = 1 caused E = 1.

$$P(m_{c=1} \rightarrow |m_{c=1}, e, a, m_{a=0}) = \frac{w_{mc1e} - w_{mc1e} w_{mae} \alpha}{w_{mc1e} + w_{mae} - w_{mc1e} w_{mae}}$$

Since M_A = 0, $w_{mc1e} w_{mae} \alpha = 0$ and the equation reduces to 1.0:

$$P(m_{c=1} \rightarrow |m_{c=1}, e, a, m_{a=0}) = \frac{w_{mc1e}}{w_{mc1e}} = 1.0$$

Step 3: multiplying the two probabilities to obtain the probability that c caused e:

$$P(c \rightarrow e|c, e, m_{c=1}, a, m_{a=0}) = P(c \rightarrow m_{c=1}|c, m_{c=1}) P(c \rightarrow m_{c=1}|c, m_{c=1}) = 1.0$$

The equations for these test cases also show that the parameter values are irrelevant in these cases.

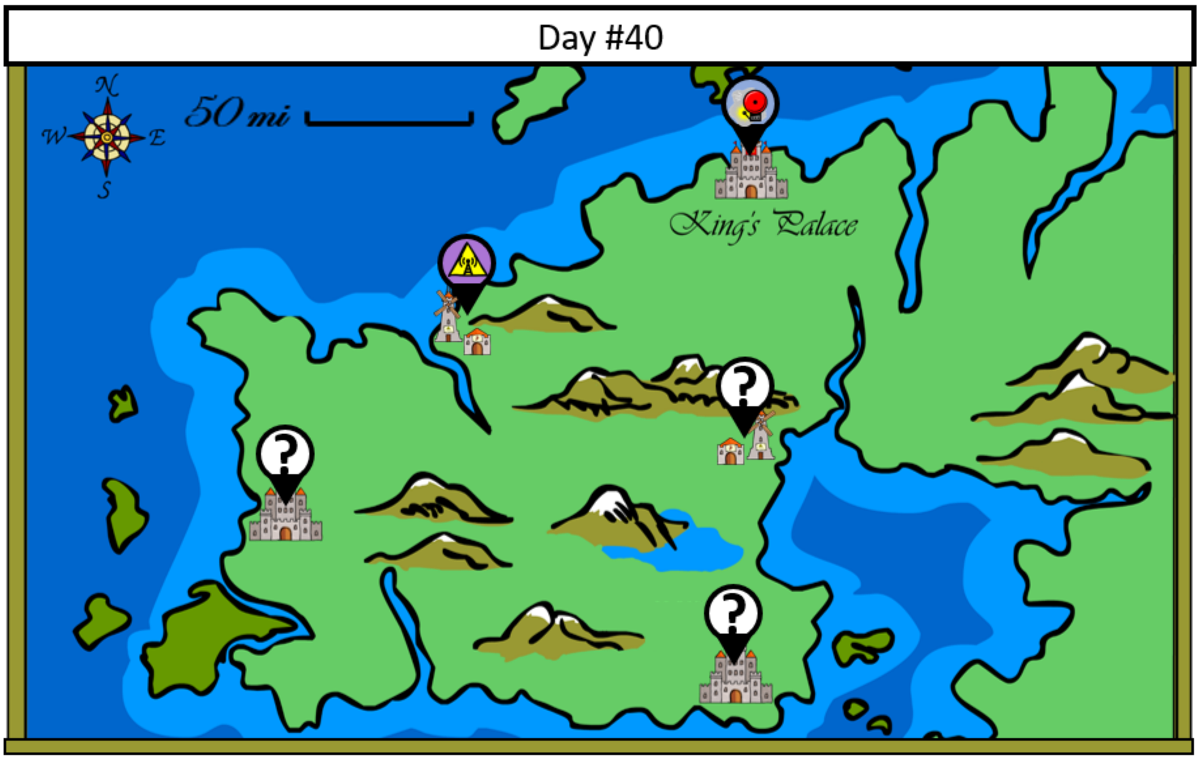
<u>Segment 5: test cases 23 – 28</u>

These cases mirror those of segment 1. Here, the target castle failed to activate its intermediate station. The results in all three steps illustrated for segment 1 thus reduce to 0.

<u>Segment 2: test cases 07 – 12</u>

For these cases, the parameter values are relevant, at least for some of the steps.

Test case 07. We see that the effect is present (an alarm occurred in the palace) and that the intermediate mechanism variable of the target castle is in its “superior” states (the intermediate station sent a telegraph). Whether the competing cause (the other castle) is active and whether its intermediate variable is active (this castle’s intermediate station) is unobserved. Whether the target cause is active is also unobserved (whether the target castle sent a pigeon). The picture below shows test case 07 in the condition in which the target cause is the Western castle.



Step 1: What is the probability that the target telegraph tower was activated by the (unobserved) Western castle?

$$P(c \to m_{c=1} | m_{c=1}) = \frac{b_c w_{cmc1} - b_c w_{cmc1} b_a w_{amc1} \alpha}{b_c w_{cmc1} + b_a w_{amc1} - b_c w_{cmc1} b_a w_{amc1}}$$

As the target telegraph tower has no alternative causes, $b_c w_{cmc1} b_a w_{amc1}$ is 0 in the equation. The equation therefore yields $P(c \to m_{c=1} | m_{c=1}) = 1$.

Step 2: What is the probability that the target telegraph tower caused the alarm in the palace? We here need to consider that the alternative castle might be active, that it might have activated its telegraph tower, or its pony rider, and that its telegraph tower or pony rider might have caused the alarm.

In this case, what is the causal strength with which the competing cause generates the effect? We need to average over the different possibilities.

$$\overline{w_a} = b_a w_{ama1} w_{ma1e} + b_a w_{ama2} w_{ma2e} - b_a w_{ama1} w_{ma1e} + b_a w_{ama2} w_{ma2e}$$

Inserting the strength values, we obtain:

$$\overline{w_a} = 0.28098$$

We also need to determine the size of the alpha parameter. The alpha parameter is one component of the probability with which the target cause is preempted by the alternative cause. It determines how likely it is that the target cause "reached" the effect earlier on occasions on which both the target and the alternative cause sufficiently strong to generate the effect. As we have shown in Stephan, Mayrhofer, and Waldmann (2020), alpha can be computed based on causal latency and onset information. In the present experimental scenario, we assume that intuitions about causal latency are conveyed by the concrete entities that we instructed, telegraphs and pony riders. We assume that participants think that telegraph towers have a shorter causal latency than pony riders: on occasions on which one station sends a telegraph and the other sends a pony, it will rarely be the case that the pony is quicker than the telegraph; only if it has a massive onset advantage. We therefore assume a very low value of 0.1 for alpha if the target mechanism variable is a telegraph and the competing mechanism variable is a pony rider. We call this advantageous state of alpha "alpha superior", α_{sup}. In the opposite case, we assume a high alpha value of 0.9. We call this disadvantageous state of alpha "alpha inferior", α_i . In cases in which both intermediate variables take on the same value (pony vs. pony or telegraph vs. telegraph), we assume that there is a 50:50 chance that the competing cause preempts the target cause. In those cases, we set alpha to 0.5. We denote this state of alpha "alpha symmetric", α_{sym}.

In test case 07, we need to average over the different possibility for alpha. The target intermediate variable is a telegraph tower. Thus, if the alternative intermediate station is also in the stage "telegraph", alpha would be 0.5. If that station is in state "pony" alpha would be 0.1. It cannot be 0.9 in this case, because the target mechanism variable is in the state "telegraph", the superior state. Finally, it is possible that the alternative station is inactive. In this case, the target cause cannot be preempted, and alpha would be 0. We now average over these possibilities to obtain the aggregate value for alpha, $\bar{\alpha}$.

$$\bar{\alpha} = b_a w_{ama1} \alpha_{sym} + b_a w_{ama2} \alpha_{sup} = 0.126$$

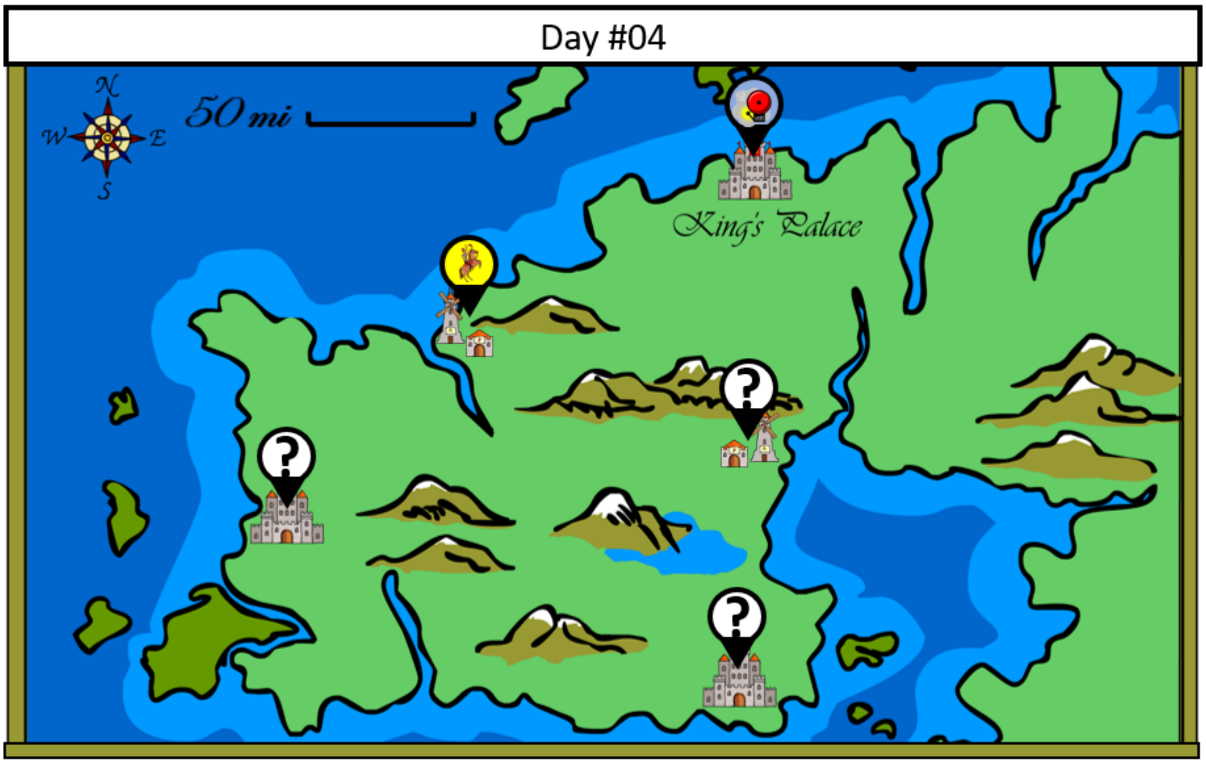
We can now insert these values in the model.

$$P(m_{c=1} \rightarrow e | m_{c=1}, e) = \frac{w_{mc1e} - \overline{w_a} \bar{\alpha}}{w_{mc1e} + \overline{w_a} - w_{mc1e} \overline{w_a}} = 0.935$$

Step 3: We can now multiply the probabilities calculated in steps 1 and 2.

$$P(c \rightarrow e | m_{c=1}, e) = P(c \rightarrow m_{c=1} | m_{c=1}) \, P(m_{c=1} \rightarrow e | m_{c=1}, e) = 0.935$$

Test case 08. The only difference from test case 07 is that the target castle this time only activated a pony rider instead of a telegraph. Thus, only step 2 is different.



Step 2:

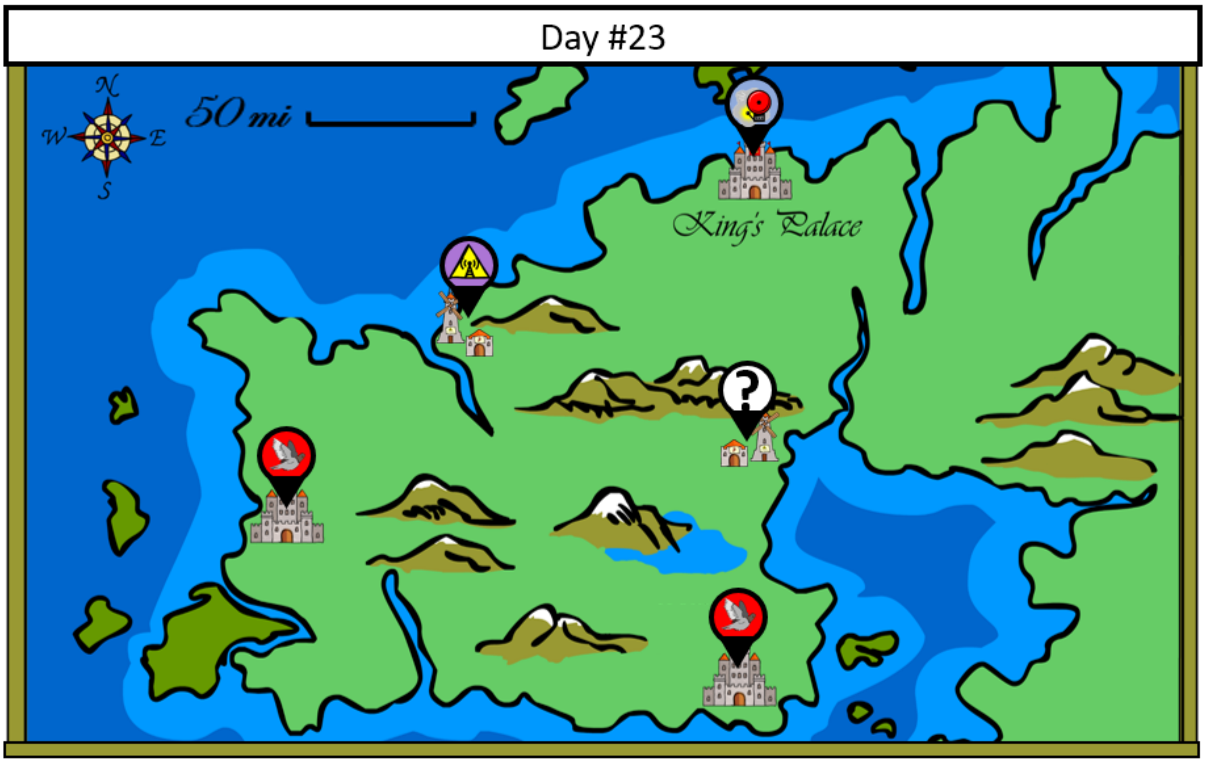
$$\bar{\alpha} = b_a w_{ama2} \alpha_{sym} + b_a w_{ama1} \alpha_i = 0.294$$

$$P(m_{c=2} \rightarrow e | m_{c=2}, e) = \frac{w_{mc2e} - \overline{w_a}\bar{\alpha}}{w_{mc2e} + \overline{w_a} - w_{mc2e}\overline{w_a}} = 0.75$$

Step3 :

$$P(c \rightarrow e | m_{c=2}, e) = P(c \rightarrow m_{c=2} | m_{c=2}) \, P(m_{c=2} \rightarrow e | m_{c=2}, e) = 0.75$$

Test case 09. The only difference from test case 07 is that we observe that the competing castle is active



.

The only difference in step 2 is that we can leave out the base rate of the competing castle.

Step 2:

$$\overline{w_a} = w_{ama1}w_{ma1e} + w_{ama2}w_{ma2e} - w_{ama1}w_{ma1e} + w_{ama2}w_{ma2e} = 0.45$$

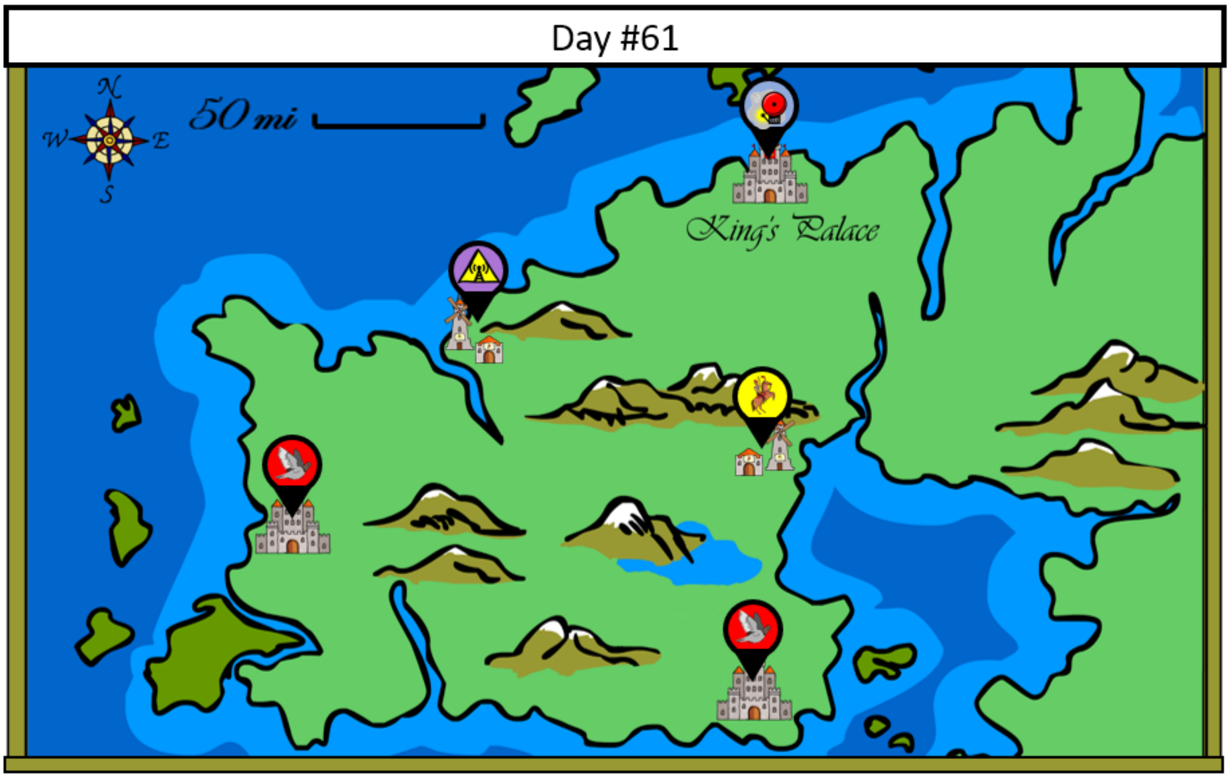
$$\bar{\alpha} = w_{ama1}\alpha_{sym} + w_{ama2}\alpha_{sup} = 0.18$$

$$P(m_{c=1} \to e|m_{c=1}, e, a) = \frac{w_{mc1e} - \overline{w_a}\bar{\alpha}}{w_{mc1e} + \overline{w_a} - w_{mc1e}\overline{w_a}} = 0.875$$

Step 3:

$$P(c \to e|c, m_{c=1}, e, a) = P(c \to m_{c=1}|m_{c=1}, c,)\, P(m_{c=1} \to e|m_{c=1}, e, a) = 0.875$$

Test case 10. The only differences from test case 07 are that we observe that the competing castle is active and that its intermediate variable is in the state “pony rider”.



In step 2, we now don’t need to compute average values for alpha and the competing cause’s strength.

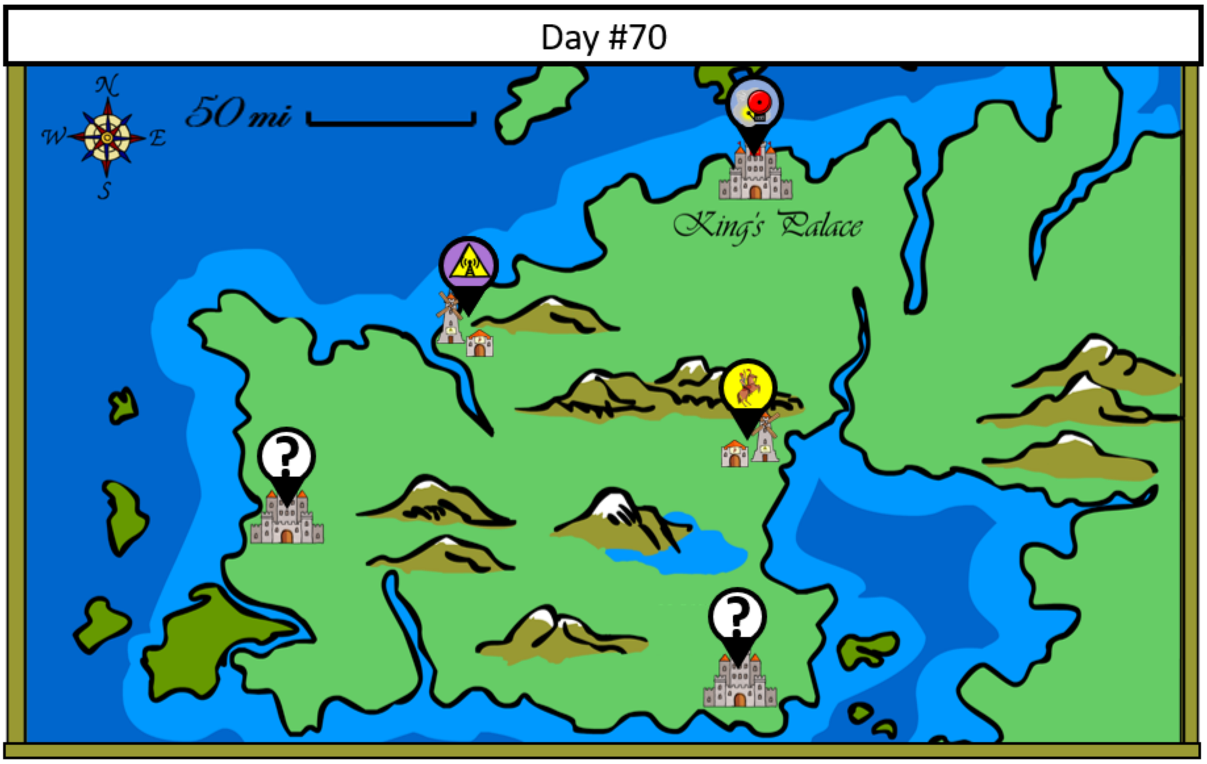
Step 2:

$$P(m_{c=1} \to e | m_{c=1}, e, m_{a=2}) = \frac{w_{mc1e} - w_{ma2e}\alpha_{sup}}{w_{mc1e} + w_{ma2e} - w_{mc1e}w_{ma2e}} = 0.881$$

Step 3:

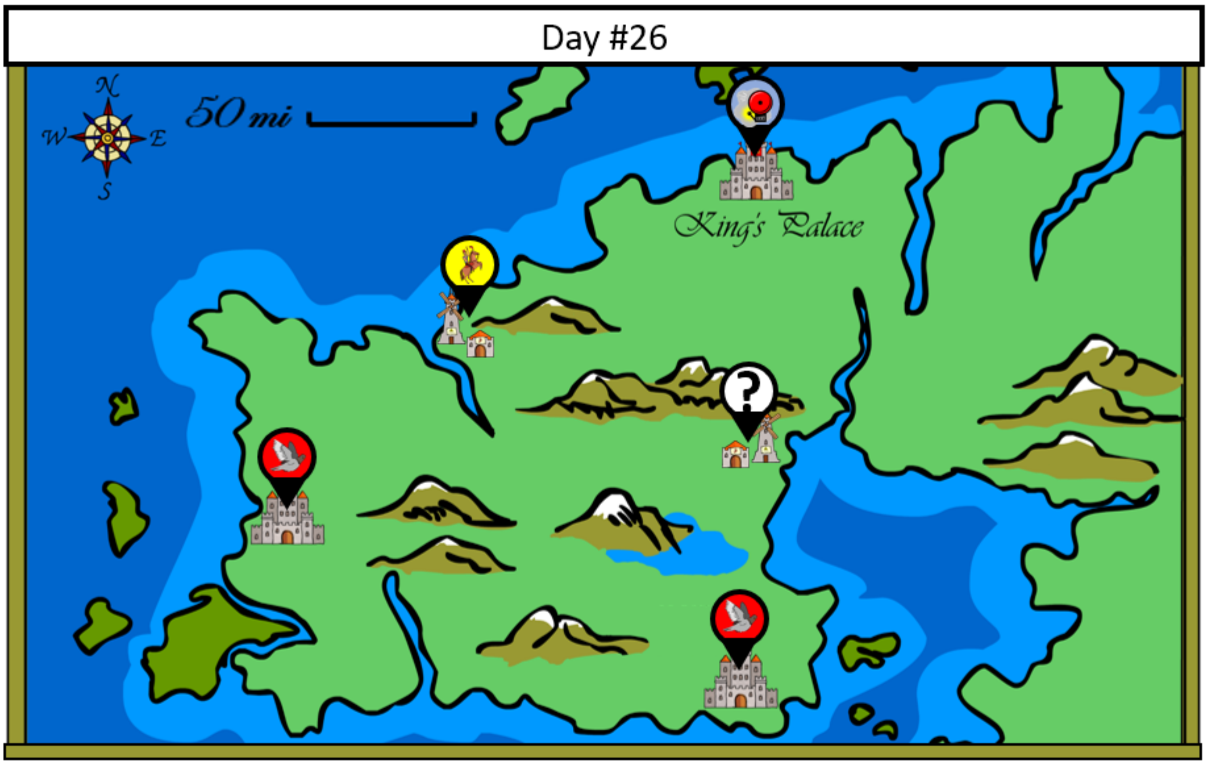
$$P(c \to e | c, m_{c=1}, e, a, m_{a=1}) = P(c \to m_{c=1} | m_{c=1}, c,)\, P(m_{c=1} \to e | m_{c=1}, e, m_{a=2}) = 0.881$$

Test case 11. The only difference in this case from the previous test case 10 is that we don't observe the states of the castles. However, as we showed the probabilities that the castles caused the present intermediate states is 1.0. We thus don't need to repeat step 1 for test case 11.



We also don't need to do step 2 again, because it is the same as for test case 10. Step 3 is thus also identical. Both cases yield the same probability of singular causation.

Test case 12. This case is similar to test case 09. The difference is that the target intermediate station this time only is in state “pony rider”. This leads to a different average value for alpha. The rest of step 2 is identical to test case 09.



Step 2:

$$\bar{\alpha} = w_{ama2}\alpha_{sym} + w_{ama1}\alpha_i = 0.42$$

$$P(m_{c=2} \to e|m_{c=2}, e, a) = \frac{w_{mc2e} - \overline{w_a}\bar{\alpha}}{w_{mc2e} + \overline{w_a} - w_{mc2e}\overline{w_a}} = 0.624$$

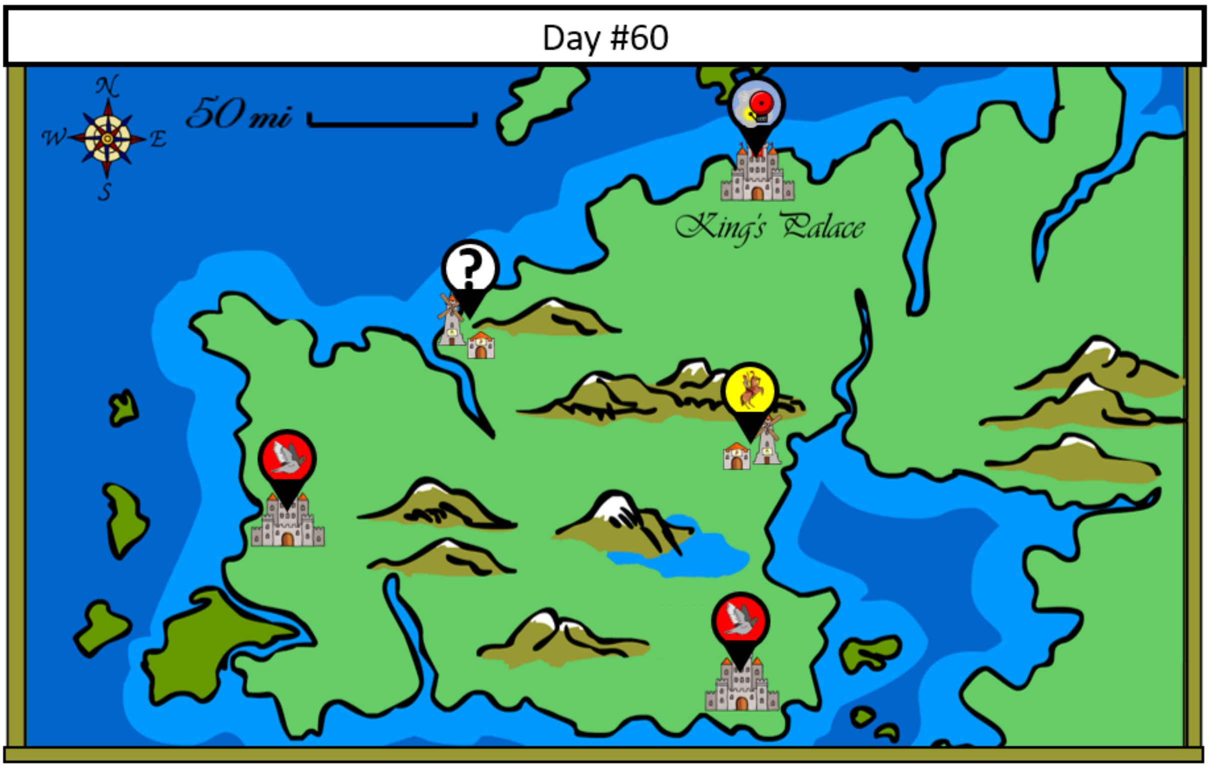
Step 3:

$$P(c \to e|c, m_{c=2}, e, a) = P(c \to m_{c=2}|m_{c=2}, c,)\, P(m_{c=2} \to e|m_{c=2}, e, a) = 0.624$$

Segment 4: test cases 17 – 22

We derive the predictions for segment 4 before segment 3 because the test cases in segment 4 mirror those of segment 2. A characteristic of our generalized model (but, for example, not of the standard power PC model of causal attribution) is that it predicts the inverse probabilities for these mirrored cases. Yet, below we derive them step by step.

Test case 17. This test cases mirrors test case 12.



To answer the question how likely it is that the alarm was caused by the target Western castle, we now first need to compute the average strength of the target castle to cause the effect in these cases. To do so, we need to average over the strengths for the different possible mechanism possibilities. We also need to average over the different possibilities for alpha.

$$\overline{w_c} = w_{cmc1}w_{mc1e} + w_{cmc2}w_{mc2e} = 0.45$$

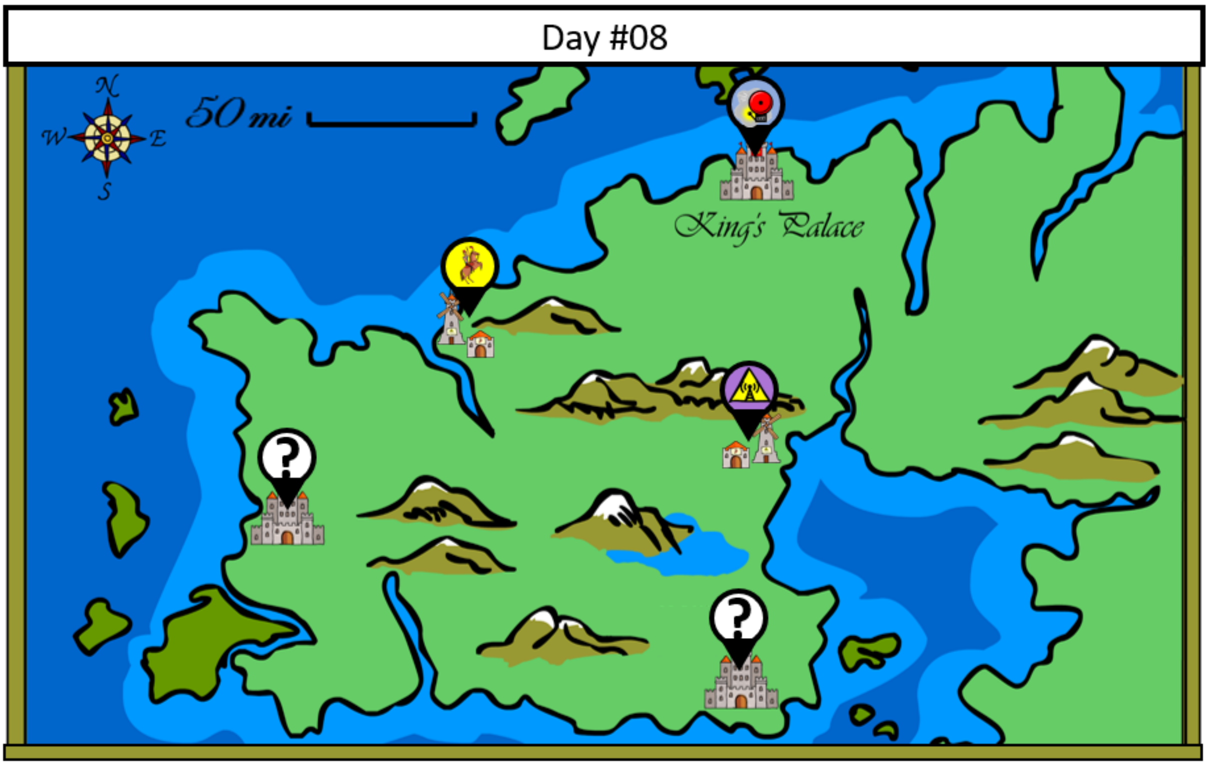
$$\bar{\alpha} = w_{cmc2}\alpha_{sym} + w_{cmc1}\alpha_{sup} + (1 - w_{cmc2} - w_{cmc1})\alpha_1 = 0.58$$

The last term in the equation for $\bar{\alpha}$ represents the possibility that the intermediate station is inactive. In this case we set alpha to 1.0.

We can now directly compute the probability that the Western castle caused the alarm.

$$P(c \to e | c, e, a, m_{a=2}) = \frac{\overline{w_c} - \overline{w_c}w_{ma2e}\bar{\alpha}}{\overline{w_c} + w_{ma2e} - \overline{w_c}w_{ma2e}} = 0.376$$

Test case 18. This test cases mirrors test case 11.

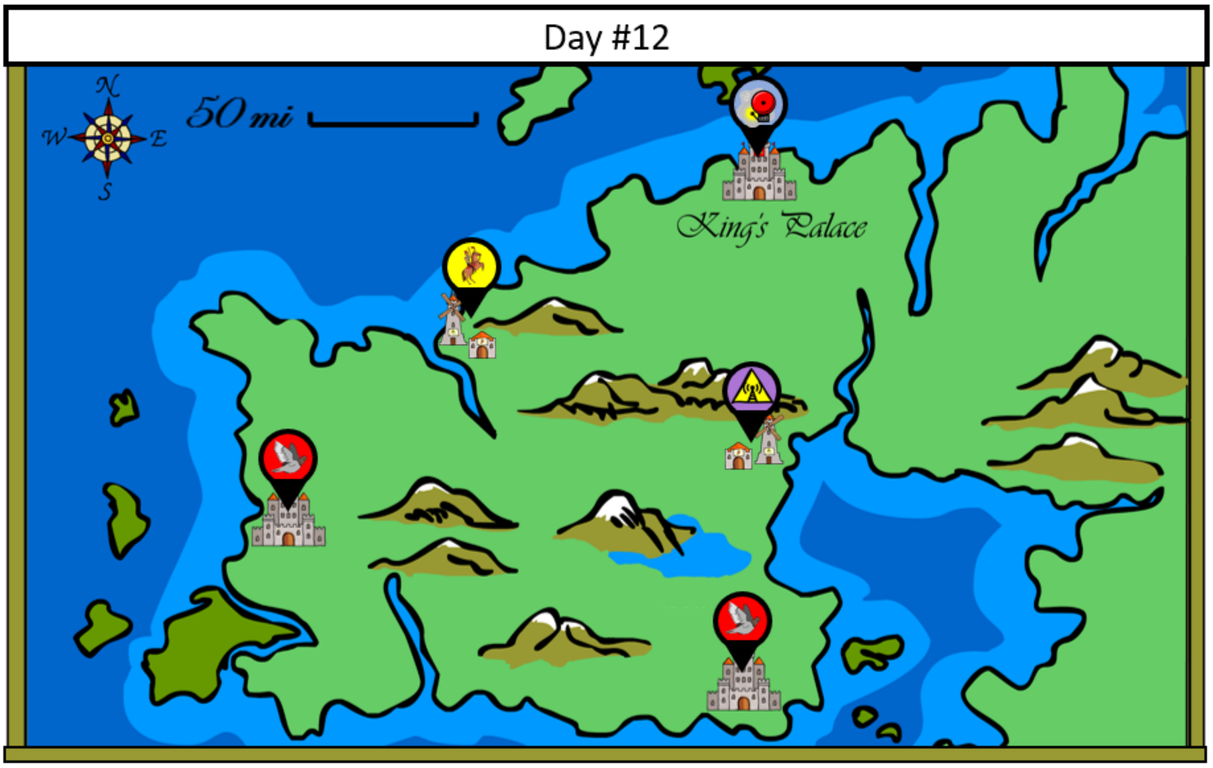


We already showed that the Western castle must have caused the "pony rider" state in the intermediate station. We therefore can directly compute the probability that the "pony state" caused the alarm.

$$P(m_{c=2} \rightarrow e | m_{c=2}, e, m_{a=1}) = \frac{w_{mc2e} - w_{ma1e}\alpha_i}{w_{mc2e} + w_{ma1e} - w_{mc2e}w_{ma1e}} = 0.11875$$

$$P(c \rightarrow e | m_{c=2}, e, m_{a=1}) = P(c \rightarrow m_{c=2} | m_{c=2})\, P(m_{c=2} \rightarrow e | m_{c=2}, e, m_{a=1}) = 0.11875$$

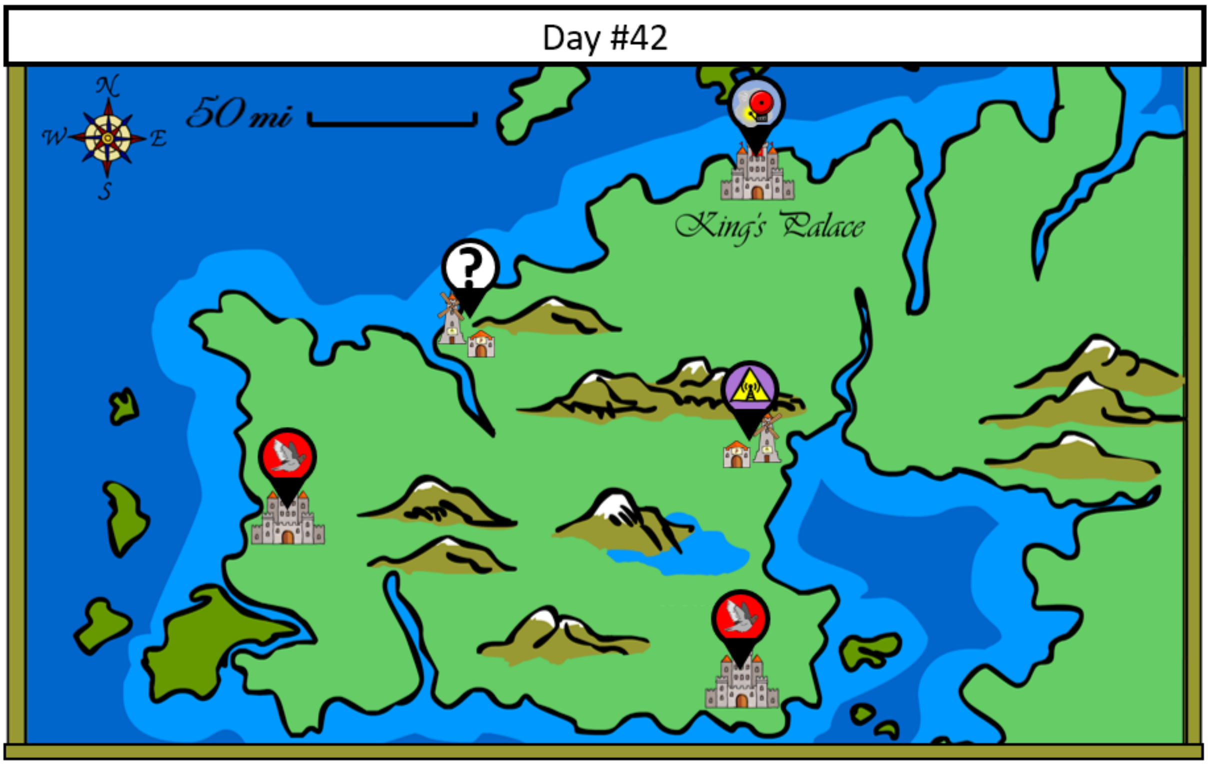
Test case 19. This test case mirrors test case 10.



The only difference from test case 18 is that we this time observe that the castles are active. However, the presence of their intermediate stations implies their presence (and that the intermediate stations can only have been caused by their castles). We therefore obtain the identical prediction as for test case 18.

$$P(c \to e|c, m_{c=2}, e, a, m_{a=1}) = P(c \to m_{c=2}|m_{c=2}, c)\, P(m_{c=2} \to e|m_{c=2}, e, m_{a=1}) = 0.11875$$

Test case 20. This test case mirror test case 09.



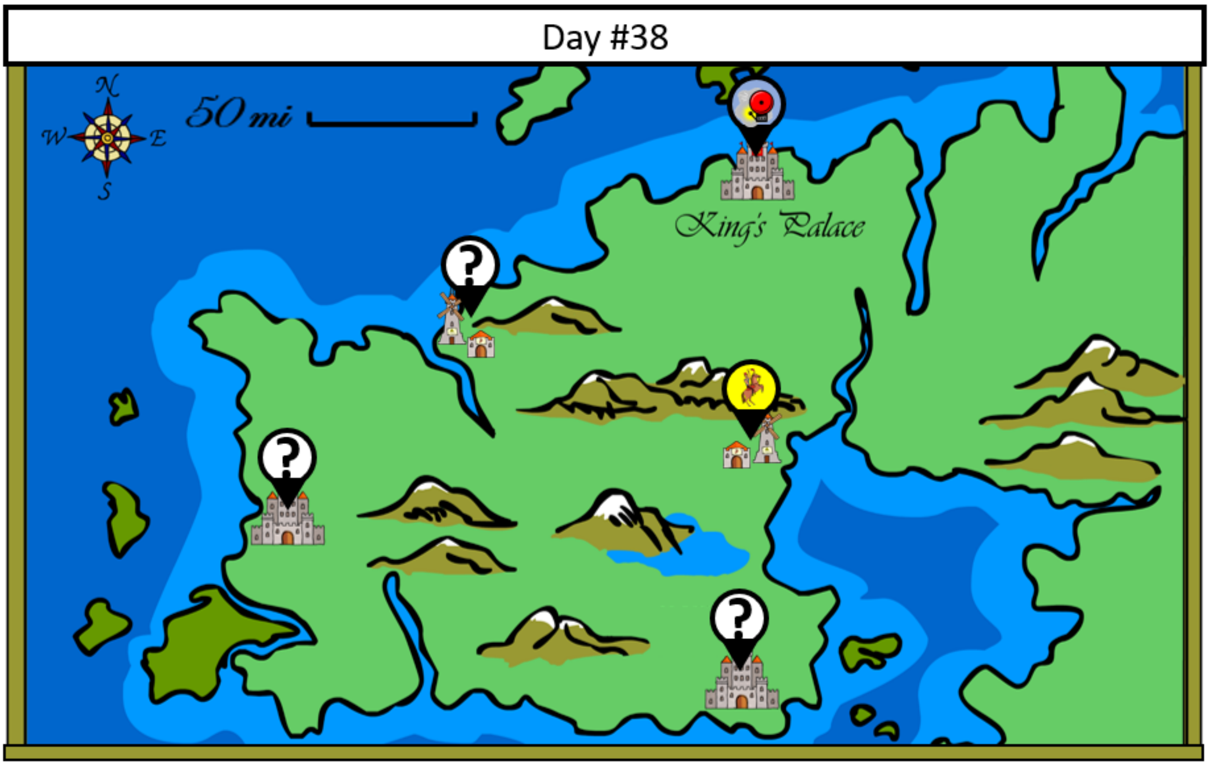
We here need to do the steps that we did for test case 17, but we need to adjust the alpha value because this time the status of the competing intermediate station is "telegraph".

$$\bar{\alpha} = w_{cmc1}\alpha_{sym} + w_{cmc2}\alpha_i + (1 - w_{cmc2} - w_{cmc1})\alpha_1 = 0.82$$

We can now directly compute the probability that the Western castle caused the alarm.

$$P(c \to e | c, e, a, m_{a=1}) = \frac{\overline{w_c} - \overline{w_c} w_{ma1e}\bar{\alpha}}{\overline{w_c} + w_{ma1e} - \overline{w_c} w_{ma1e}} = 0.125$$

Test case 21. This test case mirrors test case 08.



The probability with which the target castle causes the effect in this case needs to incorporate the castle's base rate. The same holds for the average alpha value.

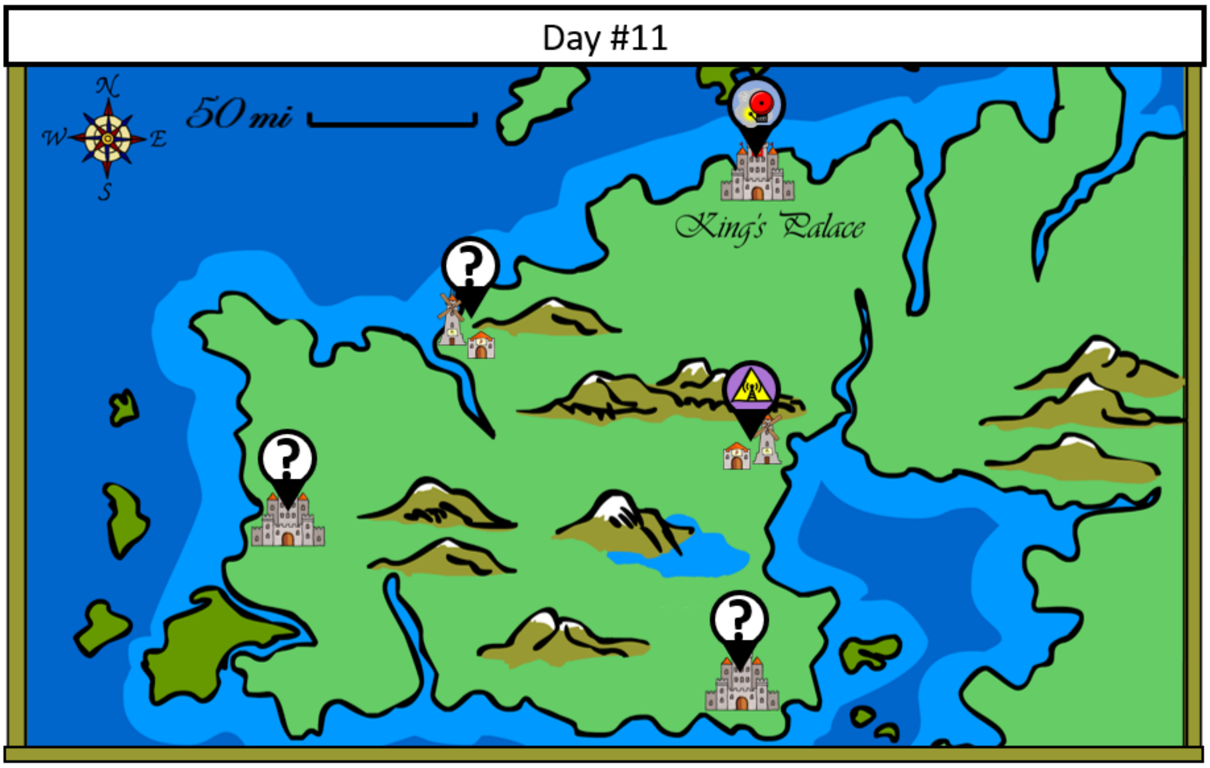
$$\overline{w_c} = b_c w_{cmc1} b_c w_{mc1e} + b_c w_{cmc2} b_c w_{mc2e} = 0.315$$

$$\bar{\alpha} = b_c w_{cmc2} \alpha_{sym} + b_c w_{cmc1} \alpha_{sup} + (1 - (b_c w_{cmc2}) - (b_c w_{cmc1}))\alpha_1 = 0.706$$

The probability that the Western castle caused the alarm is thus given by:

$$P(c \rightarrow e |\, e, a, m_{a=2}) = \frac{\overline{w_c} - \overline{w_c} w_{ma2e} \bar{\alpha}}{\overline{w_c} + w_{ma2e} - \overline{w_c} w_{ma2e}} = 0.25$$

Test case 22. This test case mirrors test case 07.



The only thing that changes in comparison to test case 21 is that alpha will be higher, because the competing intermediate station now is in state "telegraph".

$$\bar{\alpha} = b_c w_{cmc1}\alpha_{sym} + b_c w_{cmc2}\alpha_i + (1 - (b_c w_{cmc2}) - (b_c w_{cmc1}))\alpha_1 = 0.874$$

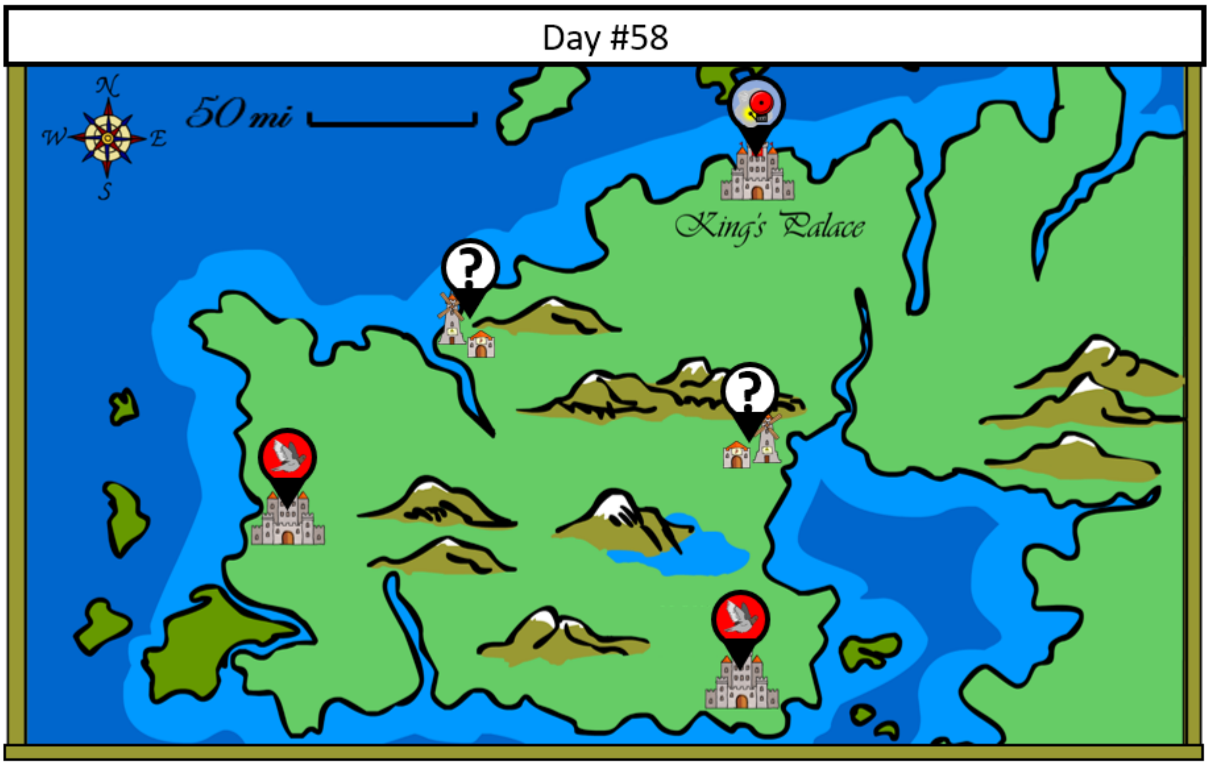
The probability that the Western castle caused the alarm is thus given by:

$$P(c \to e |\, e, a, m_{a=1}) = \frac{\overline{w_c} - \overline{w_c} w_{ma1e}\bar{\alpha}}{\overline{w_c} + w_{ma1e} - \overline{w_c} w_{ma1e}} = 0.072$$

Segment 3:

The test cases in segment 3 represent the “symmetric” test cases, for which the exactly same information is given for the two competing causes. The average alpha in these cases will always be 0.5. The average strength values for the causes will always be identical, that is, $\overline{w_c} = \overline{w_a}$. For all these cases, the probability that the target cause caused the effect will thus be 0.5. As an example, we derive the predictions for test case 13.

Test case 13.



The average strength of the Western castle to cause the alarm is:

$$\overline{w_c} = w_{cmc1}w_{mc1e} + w_{cmc2}w_{mc2e} = 0.45$$

The same holds for the Eastern castle:

$$\overline{w_a} = w_{ama1}w_{ma1e} + w_{ama2}w_{ma2e} = 0.45$$

$$P(c \to e \mid e, a) = \frac{\overline{w_c} - \overline{w_c w_a} 0.5}{\overline{w_c} + \overline{w_a} - \overline{w_c w_a}} = 0.5$$