# Structural Determinants of Interventions on Causal Systems

**Brian J. Edwards, Russell C. Burnett, and Frank C. Keil**
**Brian.Edwards@yale.edu, Russ.Burnett@yale.edu, Frank.Keil@yale.edu**
Department of Psychology, Yale University

**Abstract**

We investigate how people use causal knowledge to design interventions to affect the outcomes of causal systems. We propose that in addition to using content or mechanism knowledge to evaluate the effectiveness of interventions, people are also influenced by the abstract structural properties of a causal system. In particular, we investigated two factors that influence whether people tend to intervene proximally (on the immediate cause of an outcome of interest) or distally (on the root cause of a chain leading to the outcome). We presented people with causal chains describing a variety of real-world and artificial causal systems and asked them where they would intervene to affect the outcome. In Experiment 1, participants who were asked to choose the best long-term intervention intervened more distally than participants asked to choose the best short-term intervention. In Experiment 2, participants presented with a branching structure in which there were two distinct causal pathways from the root cause to the outcome were more likely to intervene on the root cause than participants presented with only one of the pathways. Our findings demonstrate two ways in which people integrate content knowledge and knowledge of a system's causal structure to design effective interventions.

**Keywords:** Causal reasoning; interventions; cognitive biases; knowledge structure

## Introduction

Causal knowledge is essential for understanding how the world works. Our knowledge of how causal systems work not only allows us to make predictions, it enables us to act on the world. We can use this information to design precise interventions to manipulate a causal system to achieve a desirable outcome, or alternatively, to avoid an undesirable outcome. Indeed, it has been suggested that understanding how to intervene on a system is the very essence of causal knowledge (Woodward, 2003).

For example, we can use our causal knowledge to evaluate the effectiveness of various strategies for fighting terrorism. A simplified theory of the causes of al-Qaeda terrorism might consist of the following causal chain. American meddling in the Muslim world causes anti-American attitudes, anti-American attitudes cause people to join terrorist groups, and people joining terrorist groups causes terrorism. Causal knowledge like this can be used to generate strategies, or interventions, for countering terrorism. In particular, it sheds light on potential interventions such as capturing and killing terrorists or promoting a positive image of America in the Muslim world.

Of course, our reasoning about causal systems often capitalizes on more than just our knowledge of the abstract structural properties of a system. In addition to knowing that A causes B causes C, we often have more detailed content knowledge about the variables in a causal system as well as knowledge of the mechanisms underlying these causal relationships. This knowledge often influences our reasoning about interventions; however, some reasoning about interventions may also be explained in terms of structural properties of causal systems. For example, consider a causal chain. In the causal chain $A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow ... \rightarrow A_{n-1} \rightarrow A_n$, $A_{n-1}$ is the immediate, or proximal cause of outcome $A_n$ and $A_1$ is the root, or distal cause of outcome $A_n$.

We hypothesize that:

1. Under different conditions, people prefer to intervene on root causes or more proximal causes.
2. People prefer to intervene on root causes when faced with multiple causal branches that lead to the outcome of interest.

In other contexts, psychologists have successfully explained reasoning in terms of causal structure. For example, in explaining the degrees to which different features of objects are important to membership in object categories, Ahn (1998) and Rehder and Hastie (2001) have invoked the relative positions of features in causal structure. More generally, psychologists have recently sought to account for various forms of causal learning and reasoning in terms of Bayes net theory, which is fundamentally about relative positions of variables in causal structures (e.g., Gopnik et al., 2004).

Given that causal structure has provided explanatory leverage in these other contexts, it is natural to ask how structure influences what is arguably the quintessential form of causal cognition: designing interventions. This topic has recently been addressed by Gopnik et al. (2004), Sloman (2005), and others. White (1997) proposed a "dissipation effect," whereby people judge that the effects of interventions shrink in magnitude as they propagate through a causal network. (In itself, the dissipation effect would seem to predict that people prefer proximal interventions in all circumstances.) Here we extend this general approach to understanding interventions. Although this work is exploratory, and some of our results

may be due to multiple effects, we find reliable effects of causal structure on decisions about interventions.

One effect concerns how structure might lead to different interventions when the goal is a long- versus short-term outcome. If a person is searching for a short-term solution, he or she may prefer an intervention on the immediate cause because such an intervention rapidly and directly affects the outcome, whereas an intervention on the root cause must influence every variable in the causal chain before affecting the outcome. However, if a person is searching for a long-term solution, he or she may be likely to prefer an intervention on the root cause because an intervention on the root cause addresses the underlying nature of a problem and provides a stable and permanent solution, whereas an intervention on the immediate cause might be a "quick fix" with little long-term efficacy. The theory of psychological essentialism (Medin and Ortony, 1989) claims that many entities have an underlying "essence" that causes the entities' other properties. Similarly, in causal systems in which the root cause is an essence, interventions on that cause may be especially compelling.

In the terrorism example, policy-makers searching for a short-term solution to terrorism might prefer to intervene on the immediate cause and recommend capturing and killing as many terrorists as possible; however, a long-term solution should also address the root causes of terrorism, such as aspects of American foreign policy that cause anti-Americanism, which causes people to join terrorist groups. Even if the U.S. government succeeds at killing current terrorists, new terrorists will take their place unless the government reduces anti-American sentiment in the Muslim world.

A second effect involves branching structure. Real-world causal systems are often characterized by multiple causal chains with a single root cause. That is, $A \rightarrow B \rightarrow C \rightarrow D$ and $A \rightarrow E \rightarrow F \rightarrow D$. For example, Lynch and Medin (2006) found that undergraduates' explanatory models of heart attacks typically included two distinct causal pathways leading from insufficient exercise to a heart attack. In such cases, it is possible that an intervention on the root cause would be perceived as having greater efficacy since it would affect the outcome via both causal chains. Thus, we predict that if people believe there are two causal pathways from the root cause to the outcome, they would be more likely to intervene on the root cause than people who are only aware of one of the two pathways. Similarly, when people are presented with a causal system described by a branching structure, participants' inferences about the existence of a root cause are consistent with a diversity model (Kim & Keil, 2003). That is, when participants were told that two distantly related symptoms in a branching structure had a common property, they were more likely to infer a general common root cause.

We studied two factors: (1) whether a person is seeking to affect the outcome in the short term or the long term and (2) whether a causal system has a branching structure in which there are multiple causal pathways from the root cause to the outcome. We presented individuals with causal chains and asked them which variable they would intervene on to affect the outcome. We found that participants who were asked for the best long-term intervention intervened more distally than participants who were asked for the best short-term intervention. Additionally, participants presented with two causal pathways from the root cause to the outcome were more likely to intervene on the root cause than participants presented with only one causal pathway. These results demonstrate two ways in which people's interventions can be explained in terms of the structural properties of causal systems.

## Experiment 1a

Experiment 1a tested whether framing a problem in a short-term or long-term context would influence whether people intervene proximally (near the immediate cause) or distally (near the root cause).

### Methods

**Participants:** We tested 41 adults who were recruited at busy locations on the Yale University campus in New Haven, CT. We used a between-subjects design; participants were randomly assigned to the short-term condition in which each stimulus item was presented in a short-term context ($n$ = 21) or the long-term condition in which each stimulus item was presented in a long-term context ($n$ = 20). Most participants were Yale University undergraduate students; however, other members of the Yale University and New Haven communities participated in the study. Participants received a Snapple beverage and candy bar as compensation.

**Stimuli:** Our stimuli were seven causal chains covering a range of real-world and artificial phenomena. We included artificial stimuli to account for the possibility that participants' content knowledge might influence how they intervened on real-world systems. The real-world stimuli were preventing a heart attack, improving the quality of health, caring for one's car (adapted from stimuli of Rehder & Hastie, 2001), and preventing terrorism. The artificial stimuli were a game in which aliens implant thoughts in each other's minds (based on the stimuli used by Steyvers et al., 2003), preventing a tribble from getting "Tribble Pox," and helping a sorcerer make a crystal glow in order to cast a spell. The stimuli were presented in this order, with the real-world stimuli

preceding the artificial stimuli. The causal chains varied in length from four to six variables. For each stimulus, we presented participants with a causal chain and provided them with a list of interventions that could be used to affect the outcome.

For the health and terrorism stimuli, we explicitly asked participants to select the best short-term or long-term intervention. For the heart attack and car stimuli, we used more subtle experimental manipulations. For the heart attack stimulus, the short-term and long-term manipulations were preventing a heart attack in a 70-year-old and a 30-year-old, respectively, and for the car stimulus, the experimental manipulations were preventing muffler damage to an old car and a new car, respectively. The purpose of the subtle manipulations was to test whether a short-term vs. long-term context effect depends on explicitly invoking a short-term or long-term goal. For the artificial stimuli, we used a variety of manipulations. The "health" and "alien" stimuli appear below. All stimuli can be viewed at www.yale.edu/cogdevlab/interventions. The experimental manipulations are highlighted in bold (emphasis added).

**Health Stimulus**

A government commission has identified the following causes of better health.

Better Education
↓
Access to Better Health Information
↓
Healthy Habits
↓
Better Health

Better education causes people to have access to better health information.
Access to better health information causes people to adopt healthy habits.
Healthy habits cause better health.

*The U.S. government is searching for the best **short-term [long-term]** policy to improve the quality of health in the U.S. Where in the causal chain should the U.S. government intervene to improve the quality of health in the U.S.?*

*Possible interventions*
*A. Increase education funding to improve the quality of education.*
*B. Increase access to better health information.*
*C. Encourage healthy habits.*

*The U.S. government should do intervention _____ to improve the quality of health in the U.S.*

**Alien Stimulus:**

The following information describes a sequence of events that causes Alien 4 to think "DAX."

Alien 1 Thinks "DAX"
↓
Alien 2 Thinks "DAX"
↓
Alien 3 Thinks "DAX"
↓
Alien 4 Thinks "DAX"

Alien 1 thinking "DAX" causes Alien 2 to think "DAX."
Alien 2 thinking "DAX" causes Alien 3 to think "DAX."
Alien 3 thinking "DAX" causes Alien 4 to think "DAX."
Bob has a mind zapper that can implant a thought in the mind of Alien 1, Alien 2, or Alien 3. Bob wants Alien 4 to think "DAX" at the end of the game. The game is **very short; it lasts five minutes [very long; it lasts three hours]**. Where in the causal chain should Bob intervene to cause Alien 4 to think "DAX?"

**Procedure:** After recruiting a participant, the experimenter handed the participant a pencil and paper questionnaire containing the seven stimulus items described above. The questionnaire took approximately 10 minutes to complete.

## Results and Discussion

We analyzed participants' responses based on the distance between the variable they intervened on and the outcome. That is, a participant choosing "C" in the health item was considered to have intervened one variable away from the outcome, a participant choosing "B" was considered to have intervened two variables away from the outcome, and a participant choosing "A" was considered to have intervened three variables away from the outcome. Since the lengths of the causal chains varied by item, we transformed the causal distance to a [0,1] scale with 0 being an intervention on the immediate cause and 1 being an intervention on the root cause.

We performed a repeated measures ANOVA on intervention distance with condition (short term or long term) as a between-subjects variable and causal system as a within-subjects variable to determine whether participants in one condition intervened significantly farther away from the outcome. Consistent with our hypothesis, participants in the long-term condition intervened farther away from the outcome than participants in the short-term condition ($F(1,34) = 4.27$, $p < .05$). We separately analyzed this trend for the real-world and artificial stimuli. For the four real-world stimuli, the average causal distance for interventions was significantly greater in the long-term condition than in the short-term condition ($F(1,37) = 14.1$, $p < .01$). For the three artificial stimuli, participants did not intervene significantly farther away from the outcome in either condition ($F(1,36) < 1$, n.s.). The data for the four real-world stimuli are presented in Table 1, where the variables in each causal system are shown in the order in which they formed a causal chain leading to the outcome.

Table 1: Percentage of Interventions on Each Variable

| | Condition | |
|---|---|---|
| | Short | Long |
| Unhealthy Diet | 65 | 95 |
| High Cholesterol | 10 | 0 |
| Blocked Arteries | 10 | 0 |
| Insufficient Oxygen to Heart | 15 | 5 |
| Heart Attack | | |

| | Short | Long |
|---|---|---|
| Better Education | 19 | 65 |
| Better Health Information | 24 | 25 |
| Healthy Habits | 57 | 10 |
| Better Health | | |

| | Short | Long |
|---|---|---|
| Budgeting Little Money | 10 | 10 |
| Buying Butane-Laden Fuel | 29 | 70 |
| Fuel-Filter Gasket Corrosion | 19 | 10 |
| Engine Runs Hot | 19 | 0 |
| Carbon Monoxide in Exhaust | 24 | 10 |
| Muffler Rusts | | |

| | Short | Long |
|---|---|---|
| U.S. Economic Meddling | 25 | 25 |
| Muslim Poverty | 30 | 40 |
| Wounded Pride | 10 | 10 |
| Anti-Americanism | 20 | 20 |
| People Join Al-Qaeda | 15 | 5 |
| Terrorism | | |

Although our data suggest that people intervene more distally when presented with a long-term context than with a short-term context, it is possible that our stimuli were biased such that interventions on the relatively distal variables were actually better long-term solutions, whereas interventions on the relatively proximal variables were better short-term solutions. Experiment 1b evaluated this alternative account.

For the three artificial stimuli, there were no significant differences between the average causal distance for long-term and short-term interventions. Three factors may explain the absence of an effect for these stimuli. First, it is possible that we used an ineffective manipulation to distinguish between the short-term and long-term conditions. Second, some participants may have found these stimuli obscure. Third, the artificial stimuli were rigid and contained minimal content. It is possible that some content and flexibility in the causal chain are necessary for a short-term or long-term context to affect where people intervene on the causal chain.

For the artificial stimuli, almost all participants preferred to intervene on either the immediate cause or on the root cause; less than two percent of interventions were on the intermediate cause. This shows that in the absence of content knowledge, people consider the immediate and root causes to be the best places to intervene to affect the outcome of a causal chain and scrupulously avoid the midpoint in an "endpoint bias."

## Experiment 1b

An alternative account of our results for the real-world stimuli in Experiment 1a is that we may have constructed the stimuli such that the more distal variables were better suited for long-term interventions, whereas the more proximal variables were better suited for short-term interventions. To test this hypothesis, we performed a control study in which we asked participants where they would intervene in either a short-term or a long-term context without presenting them with a causal chain. If our stimuli were biased in this manner, the data in Experiment 1b should resemble the data in Experiment 1a. If our stimuli were not biased, in Experiment 1b there should be no difference between the causal distance of participants' interventions in the short-term and long-term conditions.

### Methods

**Participants:** We tested 96 adults who did not participate in Experiment 1a. Half of the participants ($n$ = 48) were randomly assigned to the short-term condition and half of the participants ($n$ = 48) were randomly assigned to the long-term condition.

**Stimuli:** Our stimuli were the four real-world stimuli used in Experiment 1a. The stimuli were presented without information about the causal relationships between the variables; for example, participants only received the italicized portions of the health item (emphasis added).

The order of possible interventions was randomized subject to the constraint that no more than two of the interventions were in the correct position. We randomized the order in which the interventions were presented to prevent participants from constructing a plausible causal chain. In order to assess whether the order in which the interventions were presented would affect participants' responses, half of the participants received the interventions in the randomized order and half of the participants received the interventions in the reverse of the randomized order. We counterbalanced the order in which the four stimuli were presented.

**Procedure:** The procedure was identical to Exp. 1a.

### Results and Discussion

As a preliminary analysis, we performed a hierarchical regression to determine whether the order of the stimuli or the order in which the interventions were presented

influenced participants' responses and found no significant effect of stimulus order or intervention order.

We performed a repeated measures ANOVA on intervention distance with condition as a between-subjects variable and causal system as a within-subjects variable. Here, intervention distance refers to the distance of a variable from the outcome in Experiment 1a. Our analysis found no significant effect of condition on intervention distance ($F(1,90) = 3.25$ n.s.). This suggests that the effect observed in Experiment 1a cannot be attributed, at least wholly, to the more distal variables being inherently better suited for long-term solutions.

Framing a problem in a short or long-term context influenced where in a causal chain individuals intervened in order to manipulate the outcome. Consistent with our intuition that interventions on the root cause provide stable and permanent solutions to long-term problems by addressing the underlying cause and that interventions on the immediate cause provide more direct and fast solutions to short-term problems, subjects asked for the best long-term intervention intervened significantly farther back in the causal chain than subjects asked for the best short-term intervention, even if the experimental manipulation was subtle. Thus, individuals clearly incorporated their knowledge of the system's causal structure into their decision where to intervene.

## Experiment 2

Experiment 2 asked whether people would be more likely to intervene on the root cause when presented with a branching structure than when presented with a single causal chain.

### Methods

**Participants:** We tested 52 adults. Participants were recruited and compensated in the same manner as Experiment 1a. Participants were randomly assigned to the two-pathway condition ($n = 18$) or to one of two one-pathway conditions. In the first one-pathway condition ($n = 19$), participants received the first causal chain in each pair and in the second one-pathway condition ($n = 15$), participants received the second causal chain in each pair. Participants in the two-pathway condition received both causal chains.

**Stimuli:** Five pairs of causal chains were created, covering two real-world phenomena (preventing a heart attack, improving health) and the three artificial phenomena used in Experiment 1a. Participants in the two-pathway condition were presented with both causal chains in each pair and participants in the one-pathway conditions were presented with only one of the causal chains in each pair. 18 of the 52 participants received two additional real-world stimuli covering employee productivity and road rage. The order of the stimuli was heart attack, health, aliens, tribbles, crystals, employee productivity, and road rage.

**Procedure:** The procedure was identical to Exp. 1a.

### Results and Discussion

As a preliminary analysis, we performed a repeated measures ANOVA on intervention distance with condition (first one-pathway or second one-pathway) as a between-subjects variable and causal system as a within-subjects variable. We found no difference across the two one-pathway conditions ($F(1,30) < 1$, n.s.).[1] Thus, we collapsed these two conditions in all subsequent analyses.

A repeated measures ANOVA on intervention distance with condition (one pathway or two pathways) as a between-subjects variable and causal system as a within-subjects variable found a main effect of condition ($F(1,47) = 20.9$, $p < .001$). In particular, participants in the two-pathway condition were more likely to intervene on the root cause than participants in the one-pathway condition. In the one-pathway condition, participants intervened on the immediate cause 21 times (23% of interventions) and intervened on the root cause 57 times (63%). In the two-pathway condition, participants intervened on an immediate cause four times (9%) and intervened on the root cause 38 times (83%). This effect was especially strong for the three artificial stimuli. In the one-pathway condition, participants intervened on the immediate cause 64 times (63% of interventions) and intervened on the root cause 36 times (35%). In the two-pathway condition, participants intervened on an immediate cause 13 times (24%) and intervened on the root cause 41 times (76%). As in Experiment 1a, participants overwhelmingly intervened on either the immediate cause or the root cause; less than two percent of interventions were on the intermediate cause.

The magnitude of the effect may have been stronger for the artificial stimuli than for the real-world stimuli because participants' content knowledge may have influenced how they intervened on real-world systems. For example, if a person believes that increasing access to healthcare is the best way to promote better health, he or she may choose the intervention "increase healthcare funding to increase access to healthcare," regardless of its location in the causal system. Although knowledge of the existence of multiple causal pathways with a common root cause makes people more likely to intervene on the root cause, this trend is diminished by content knowledge.

[1] The ANOVAs did not include the employee productivity and road rage stimuli, as the sample sizes for these stimuli were small. The differences were in the predicted direction.

## General Discussion

Causal structure, in addition to content knowledge, seems to influence evaluations of the effectiveness of alternative interventions. For one thing, people show a tendency to intervene on either the immediate cause or the root cause, but prefer to avoid intervening on intermediate causes. Furthermore, people seeking a long-term solution prefer to intervene towards the beginning of a causal chain, especially on the root cause, when compared to people seeking a short-term solution, who show a greater preference to intervene towards the end of a causal chain, especially on the immediate cause. Additionally, when people know that a causal system has a branching structure with multiple routes to the outcome, they show a greater tendency to intervene on the root cause.

Our research may be of interest to policy-makers who seek to persuade others of the value of a particular intervention. Policy-makers advocating an intervention on the immediate cause should frame the problem in a short-term context, whereas policy-makers advocating an intervention on the root cause should present the problem in a long-term context. In the latter case, it may also be useful to suggest multiple causal chains from the root cause to the outcome.

This work raises questions concerning the development of intuitions about interventions. Do children consider causal structure when deciding where to intervene on a causal chain? Children's understanding of links between causal knowledge and interventions has been studied in the context of causal learning (see Schulz, Kushnir, & Gopnik (2007) for a review); however, little research has addressed how children use their causal knowledge to manipulate the outcomes of causal systems. Studies examining children's intuitions about such interventions can provide cues into whether the sensitivity to structure observed in adults' interventions emerges early in development or only arises after extensive experience.

Future studies might also ask whether our results are the result of a general tendency or the composition of several smaller effects. For example, an intervention on the immediate cause is likely to affect the outcome faster and more reliably than an intervention on the root cause. Would this effect persist in cases where interventions on the root cause and interventions on the immediate cause affect the outcome with equal speed and reliability? Interventions on the root cause are also attractive because they affect all variables in the causal chain; thus, intervening on the root cause might have desirable effects on variables besides the outcome. In the terrorism stimulus, people might want to intervene to reduce American meddling because it will improve America's image in addition to reducing terrorism. In a causal chain in which the intermediate variables have neutral value, there may be less of a tendency to intervene on the root cause. Finally, interventions on root causes in a single chain might be preferred because of inferred secondary routes. The effect might shrink if such routes are blocked.

Whether one is seeking to influence an intervention decision or simply optimize reasoning about where to intervene, these studies suggest that it may be important to take into account cognitive trends arising from causal structure.

## Acknowledgements

We thank Greg Laun for assistance with data analysis, five anonymous reviewers for valuable suggestions, and NIH Grant R37HD023922 to Frank C. Keil.

## References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. Cognition 69, 135-178.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review, 111, 3-32.

Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. Memory and Cognition, 31(1), 155-165.

Lynch, E., & Medin, D. (2006). Explanatory models of illness: A study of within-culture variation. Cognitive Psychology, 53(4), 285-309.

Medin, D.L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 179–196). Cambridge, MA: Cambridge University Press.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction and similarity. Journal of Experimental Psychology: General, 130, 323-360.

Schulz, L. E., Kushnir, T., & Gopnik, A. (2007). Learning from doing: Interventions and causal inference. In A. Gopnik & L. E. Schulz (Eds.), Causal Learning; Psychology, Philosophy, and Computation. New York: Oxford University Press.

Sloman, S. (2005). Causal models: How people think about the world and its alternatives. New York: Oxford University Press.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E., & Blum, B. (2003). Inferring causal networks from observations and interventions. Cognitive Science, 27, 453-389.

White, P. A. (1997). Naive ecology: Causal judgments about a simple ecosystem. *British Journal of Psychology, 88,* 219–233.

Woodward, J. (2003). Making things happen: A theory of causal explanation. New York: Oxford University Press.