# Causal Structure in Conditional Reasoning

**Tevye R. Krynski (tevye@mit.edu)**
**Joshua B. Tenenbaum (jbt@mit.edu)**
Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139

**Abstract**

Causal reasoning has been shown to underlie many aspects of everyday judgment and decision-making. We explore the role of causal structure in conditional reasoning, hypothesizing that people often interpret conditional statements as assertions about causal structure. We argue that responses on the Wason selection task reflect the selection of evidence expected to maximally reduce uncertainty over candidate causal structures. We present a model in which people's selections depend on their interpretation of which causal relationship is asserted by a given conditional statement.

## Introduction

Consider the following statement: "If a pot falls in the kitchen, then you will hear a clang". Is this statement true? Not if something breaks its fall, like a pillow. Now consider the statement: "If a clang is heard then a pot has fallen in the kitchen." Is this statement true? Not if something else can cause a clang, such as falling silverware. The first statement is not always true because there are conditions that can disable the mechanism by which falling pots cause clangs to be heard. The second statement is not always true because there are alternate causes of clangs other than falling pots. As this example illustrates, causal knowledge often underlies how people reason about conditional statements.

Recent research has shown that causal reasoning permeates many aspects of cognition, including associative learning (Waldmann, 2000; Glymour & Cheng, 1998), category learning (Rehder, 2003; Ahn, 1999), and judgment under uncertainty (Krynski & Tenenbaum, 2003). In this paper we analyze the role of causal structure in conditional reasoning (Over & Jessop, 1998), and argue that people's responses on the Wason selection task reflect sophisticated abilities to induce causal structure.

An important open question in causal reasoning is how people's background knowledge interacts with observations when inferring causal structure. Causal domain knowledge places important constraints on which cause-effect relationships exist and how the effects depend functionally on the causes (Pearl, 2000; Krynski & Tenenbaum, 2003; Ahn, Kalish, Medin, & Gelman, 1995). This effectively specifies a hypothesis space of candidate causal structures, which we model using causal Bayes nets (Pearl, 2000). Observational evidence can then be used to determine which causal structure is most likely. We propose that this interplay of causal domain knowledge and observational evidence underlies people's judgments on the Wason selection task.

The Wason selection task presents subjects with a conditional statement of the form "if $p$ then $q$", and asks subjects to choose evidence to determine whether the statement is true. Prior accounts of people's responses on the selection task have emphasized logical reasoning (Wason, 1966; Ahn & Graham, 1999), probabilistic reasoning (Oaksford & Chater, 1994), or social reasoning (Cosmides, 1989), as well as others. In contrast, we argue that the selection task often engages causal reasoning: for conditional statements in which $p$ and $q$ are causally related, people choose cards that will be most useful to determine which of several candidate causal structures is correct for a given situation.

We have developed a model that extends Oaksford & Chater's (1994) probabilistic information gain framework to handle causal hypotheses. The information gain framework of O&C proposes that in the Wason selection task, people seek to reduce their uncertainty among hypotheses about the relationship between the antecedent ($p$) and the consequent ($q$) in a conditional statement of the form "if $p$ then $q$". The model of O&C (1994) proposes that these hypotheses are assertions about conditional dependencies (e.g., $q$ depends on $p$, $q$ is independent of $p$, etc.), whereas we propose that these hypotheses are assertions about causal structure (e.g., $p$ causes $q$, $p$ does not cause $q$, etc.).

Our causal framework enables us to explain some previously puzzling results from the literature, as well as compelling intuitions that are not predicted by other approaches. We also address an important open question with both logical and probabilistic accounts: they leave unspecified how people interpret conditionals to determine which hypothesis is being asserted. We propose that the interpretation of conditionals often depends on causal domain knowledge, which imposes constraints on candidate causal structures, as well as pragmatic considerations.

## Why interpret conditionals causally?

In contrast to O&C's proposal that conditional statements assert a conditional dependency, we propose that people interpret conditional statements in which $p$ and $q$ are causally related as assertions about causal structure. The underlying reason for this is that conditional dependencies are often a symptom of some underlying causal relationship. "If $p$ then $q$" states that there is some dependency between $p$ and $q$, which in turn implies there is some mechanism by which $p$ and $q$ are related; i.e., $p$ causally influences $q$, $q$ causally influences $p$, or they have some common cause. The term "causally influences" does not necessarily mean

"directly causes"; causal influence can be generative, inhibitive, enabling, permissive, or otherwise.

Examples of the prevalence of causal interpretations come from statements in which the logical interpretation and the causal interpretation are at odds; in these cases, the causal interpretation tends to take precedence. Some conditionals are logically false but seem true because they are causally true. For example, "If you spin around then you will get dizzy" seems true enough, although it's possible to spin around without getting dizzy, therefore it's logically false. Other conditionals are logically true but seem false because they are causally false. For example, "If you drink coffee during the day then you will fall asleep at night" sounds false because it seems to be saying that coffee causes you to fall asleep, but it is logically true (assuming you eventually fall asleep every night). These examples suggest that it is often, but not always, more natural to interpret conditionals as causal assertions, rather than logical implications.

## Causal Structure Induction

We adopt the following Bayesian framework: given conditional statement "if $p$ then $q$", reasoners consider a total hypothesis space $T$ of candidate causal structures relating $p$ and $q$. The conditional statement is interpreted to be asserting that a specific causal relationship holds between $p$ and $q$. $T$ then partitions into a subspace of structures $S$ consistent with the statement, and its complement, $T$-$S$, inconsistent with the statement. Testing the conditional amounts to testing whether the true structure, $s^*$, is in $S$ or $T$-$S$. The probability that the conditional is true is the probability that $s^*$ is in $S$: $P(s^* \in S) = P(S) = \sum_{s \in S} P(s)$.

Initial degrees of belief in these hypotheses are represented as prior probabilities, and those structures that do not satisfy the constraints of causal domain knowledge are not considered. For example, people know that falling pots can cause noise, but noise cannot cause pots to fall, hence no structures with noise causing falling pots will be included in $T$. In this case of the conditional "if a pot falls, then it makes a noise", $T$ could be the set of all causal structures consistent with domain knowledge in which falling pots exist, and $S$ could be the subset of structures in $T$ in which falling pots are a cause of noise.

Data can help determine how likely the conditional statement is to be true. Using Bayesian belief updating,

$$P(S \mid d) = \sum_{s \in S} P(s \mid d) = \sum_{s \in S} \frac{P(s)P(d \mid s)}{P(d)}$$

According to the information gain (IG) approach (O&C, 1994), when determining whether a particular conditional statement is true, the most informative data are those that are expected to maximize information gain, $I_g$:

$$I_g(S \mid D) = \sum_H P(H) \log \frac{1}{P(H)} - \sum_H P(H \mid d) \log \frac{1}{P(H \mid d)}$$

However, O&C (1996) propose that when $P(S)$ is not 0.5, a better measure is the distance between the probability distributions of the new and old beliefs, as measured by Kullback-Leibler distance, (we will use this, and call it $I_{KL}$):

$$I_{KL}(S \mid d) = P(S \mid d) \log\left(\frac{P(S \mid d)}{P(S)}\right) + P(T-S \mid d) \log\left(\frac{P(T-S \mid d)}{P(T-S)}\right)$$

In the case of the Wason selection task, $I_{KL}(S|d)$ is the amount of information gained from turning over cards. The selection task can be used to test two claims: (1) people often interpret conditional statements in which $p$ and $q$ are causally related as assertions that a particular causal relationship holds, and (2) people select information with the goal of maximally reducing uncertainty in that assertion.

### Applying the IG approach to the selection task

The Wason selection task and its variants present people with a conditional statement of the form "if $p$ then $q$", where $p$ and $q$ can be any propositions. Cards are then presented which represent trials; one side specifies whether $p$ was true on the trial, while the other side specifies whether $q$ was true. Subjects are presented with four cards, having each of the four possible sides ($p$, $q$, $\neg p$, $\neg q$) facing up. The specific task instructions vary depending on the experimenter's intent, but they generally instruct participants to select only those cards necessary to turn over in order to determine whether or not the given conditional statement is true.

Consider the information gained from turning over a single card with $v$ on the visible side and finding $u$ on unseen side: ($v$,$u$ take on values in $\{p, q, \neg p, \neg q\}$, subject to the constraints of the selection task):

$$I_{KL}(S \mid d) = I_{KL}(S \mid v,u) = P(S \mid v,u) \log\left(\frac{P(S \mid v,u)}{P(S)}\right) + P(T-S \mid v,u) \log\left(\frac{P(T-S \mid v,u)}{P(T-S)}\right)$$

$$P(S \mid v,u) = \sum_{s \in S} P(s \mid v,u) = \sum_{s \in S} \frac{P(s \mid v)P(u \mid v,s)}{P(u \mid v)}$$

Since it is generally obvious that the cards in the Wason selection task were not randomly sampled, but rather one card of each possible side ($p$, $q$, $\neg p$, or $\neg q$) was presented, no information can be gained from learning that the visible side of the card is $v$, thus $P(s \mid v) = P(s)$.

One more step is necessary for predicting card selection: summing over all possible values of the unseen side of the card to obtain the expected information gain from turning the card with $v$ on the visible side, $EI_g(S,v)$:

$$EI_g(S,v) = \sum_u I_{KL}(S \mid v,u) P(u \mid v)$$

The IG approach proposes that subjects select cards in the Wason selection task as a function of expected information gain, with selection favoring cards with higher expected information gain.

### Applying the IG approach to causal hypotheses

The Bayesian framework presented thus far is similar to O&C (1996), except that it treats conditional statements as asserting the validity of a set of hypotheses rather than a single hypothesis. We now turn to the major differences between our account and that of O&C:

(1) The hypotheses in our framework are assertions about causal structure rather than conditional dependency; hence, a causal framework predicts different values for information gain than do O&C.
(2) We propose that mapping conditional statements onto hypotheses about causal structure is inherently ambiguous and depends on pragmatic considerations.

Here we will discuss the implications of (1), leaving the implications of (2) for the next section.

The information gained from turning over card *v* and finding *u* on the other side depends on the hypotheses under consideration; in particular, $D_{KL}(S \mid v,u)$ depends on $P(v \mid u,h)$ for every $h \in T$, which in turn depends on the content of each hypothesis *h*. In the O&C (1994) approach, *H* is the hypothesis that *q* depends deterministically on *p*, while $\neg H$ is the hypothesis that *p* and *q* are independent, and these are the only two hypotheses considered. Thus,

$$P(q \mid p,H)=1; P(q \mid \neg p,H)=b; P(p \mid q,H)=a/b; P(p \mid \neg q,H)=0$$
$$P(q \mid p,\neg H)=P(q \mid \neg p,\neg H)=b; P(p \mid q,\neg H)=P(p \mid \neg q,\neg H)=a$$

where the parameters $a=P(p)$ and $b=P(q \mid \neg p)$ are the same for *H* and $\neg H$. Other possible hypotheses are proposed by O&C but not developed, specifically those in which *q* depends probabilistically on *p*, such that $P(q \mid p,H)<1$.

In our approach, the hypothesis space *T* consists of causal structures. The conditional statement asserts that a particular causal relationship holds between *p* and *q*, thus the true causal structure is in the set *S* of structures for which this relationship holds ($S \subset T$). For a given causal structure, *h*, $P(u \mid v,h)$ can be derived using the formalism of causal Bayes nets (Pearl, 2000). For the subsequent presentation we will work with a simple causal structure that provides a reasonable approximation to many of the causal structures asserted by common conditional statements. In this structure, a cause (*C*) generates an effect (*E*), but there are conditions (*D*) that can disable the mechanism, and there are alternative causes (*A*) of the effect (see Figure 1). *D* represents all disabling conditions aggregated together, and *A* represents all alternative causes aggregated together. The arrow coming from *D* in Figure 1 indicates that the presence of *D* blocks the causal path from *C* to *E*. This structure is the causal model behind Cheng's power-pc theory (Cheng, 1997) (where $P(\neg D)$ is equal to the causal power of *C* to generate *E*); the model can also be expressed as a noisy-or Bayes net (Glymour & Cheng, 1998). For this simplified structure, the total hypothesis space *T* contains all structures with one or more of the links shown in Figure 1 (subject to the constraint that the link from *D* cannot exist without the link from *C* to *E*).

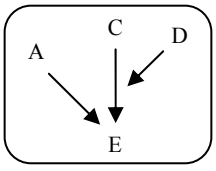


Figure 1: Noisy-or causal model

As an example of the model of Figure 1, consider a dropped pot (*C*) causing a clang (*E*). This can be disabled by various things (*D*), such as someone catching the pot or a pillow breaking the pot's fall. There are also alternate causes (*A*) of clangs, such as falling silverware.

Next we will use the semantics of the noisy-or Bayes net to derive $P(u \mid v,h)$, for the case where *h* is the hypothesis that the model of Figure 1 holds. This derivation works for all cases in which the cards in the selection task contain *C* on one side and *E* on the other, as is the case in our example "if a pot is dropped then a sound is heard" (here *p* is *C* ("a pot is dropped") and *q* is *E* ("a sound is heard"), hence *v, u* take on values in {*C, E,* $\neg C$, $\neg E$}):

$$P(E \mid C,h)=P(\neg D \vee A \mid h)=P(\neg D)+P(D)P(A)$$
$$P(E \mid \neg C,h)=P(A)$$
$$P(C \mid E,h)=\frac{P(C)P(E \mid C,h)}{P(E \mid h)}=\frac{P(C)\big(P(\neg D)+P(D)P(A)\big)}{P(A)+P(C)P(\neg D)P(\neg A)}$$
$$P(C \mid \neg E,h)=\frac{P(C)P(\neg E \mid C,h)}{P(\neg E \mid h)}=\frac{P(C)P(D)P(\neg A)}{P(\neg A)\big(P(\neg C)+P(C)P(D)\big)}$$

where the prior probabilities of *C* and *A*, *P*(*C*) and *P*(*A*), correspond to *a* and *b* in the O&C model, and the prior probability of *D*, *P*(*D*), is 1-*P*(*E*|*C*). (*P*(*D*) is taken to be zero in O&C's model for the case where *p* is *C* and *q* is *E*.) We do not require that the parameter values be the same across hypotheses, eliminating some objections to O&C's model. One could, for example, interpret a statement to be asserting that alternate causes are rare, hence *S* is all structures with P(*A*)<0.1. For simplicity, however, we will discuss only those interpretations in which a structural claim is being made (such as, a link exists from *A* to *E*); for these cases, the parameters will be the same across hypotheses.

## Interpreting conditionals as causal assertions

Conditional statements are inherently ambiguous. Those for which *p* could be a cause of *q* we will call "forward" conditionals. They generally assert that *p causes q*, but the exact causal structure being asserted depends on pragmatics. For example, the statement "if a pot is dropped then it makes a clang" could have several different meanings, as demonstrated by the following hypothetical exchanges:

(1) A: "What sound will be made if I drop this pot?"
B: "If a pot is dropped then it makes a clang."
*Meaning: dropped pots cause clangs*
*Causal Assertion:* ***dropped pots*** *can cause* ***clangs***
*Hypothesis Space: all structures in which* ***dropped pot*** *is the cause and a* ***sound*** *is the effect*

(2) A: "I think a pot just fell."
B: "That's impossible; I didn't hear a clang. If a pot falls then it makes a clang."
*Meaning: dropped pots always cause clangs*
*Causal Assertion: no D exists to block the path from* ***dropped pots*** *to* ***clangs***
*Hypothesis Space: all structures in which* ***dropped pot*** *is a cause of* ***clangs***

In contrast, conditionals for which *q* could causally influence *p* we call "reverse" conditionals. They generally assert that *q is the only cause of p*, but again the exact causal structure being asserted depends on pragmatics. For example, the statement "if you hear a clang then a pot was dropped" could have several different meanings, as

demonstrated by the following hypothetical exchanges:

(1) A: "What are those sounds coming from the kitchen?"
B: "Those are items being dropped. For instance, if you hear a clang then a pot was dropped."
*Meaning: falling pots are the primary cause of clangs, but not necessarily the only possible cause.*
*Causal Assertion:* ***dropped pots*** *can cause* ***clangs***
*Hypothesis Space: all structures in which* ***dropped pot*** *is the cause and a* ***sound*** *is the effect.*

(2) A: "I heard a clang. What do you think happened?"
B: "It must have been a dropped pot. If you hear a clang then a pot was dropped."
*Meaning: the only cause of a clang is a dropped pot.*
*Causal Assertion: no alternative cause A exists that can cause* ***clangs****.*
*Hypothesis Space: all structures in which* ***dropped pot*** *is the cause and* ***clang*** *is the effect.*

## Predicting card selection

The key point of distinction between our model and that of O&C is in predicting information gain, because $EI_g$ is a simple function of information gain. $I_{KL}(S|v,u)$ depends on the particular set of causal structures in the hypothesis space $T$, as well as the set of asserted hypotheses $S$, and the parameters $P(C)$, $P(A)$, and $P(D)$. $S$ in turn depends on pragmatic considerations. In Figures 2 and 3 we give qualitative predictions for $S$, $T$, and $I_{KL}$ for common types of conditionals, which will motivate future experiments.

Some generalizations are worth noting: (1) $EI_g(p)$ is often high. (2) for rare $p$ and $q$, when structures with no $C$ to $E$ link have sufficient priors (e.g., a uniform prior), $EI_g(S,q)$ is often high; (3) if the conditional assumes the $C$ to $E$ link exists, then structures with no $C$ to $E$ link will have low priors and $EI_g(S,q)$ will be low. In general, with pragmatic considerations, the model predicts selection of $p$ and $q$ cards if the conditional *asserts* that $C$ causes $E$, and selection of $p$, $\neg q$ if the conditional *assumes* that $C$ causes $E$.

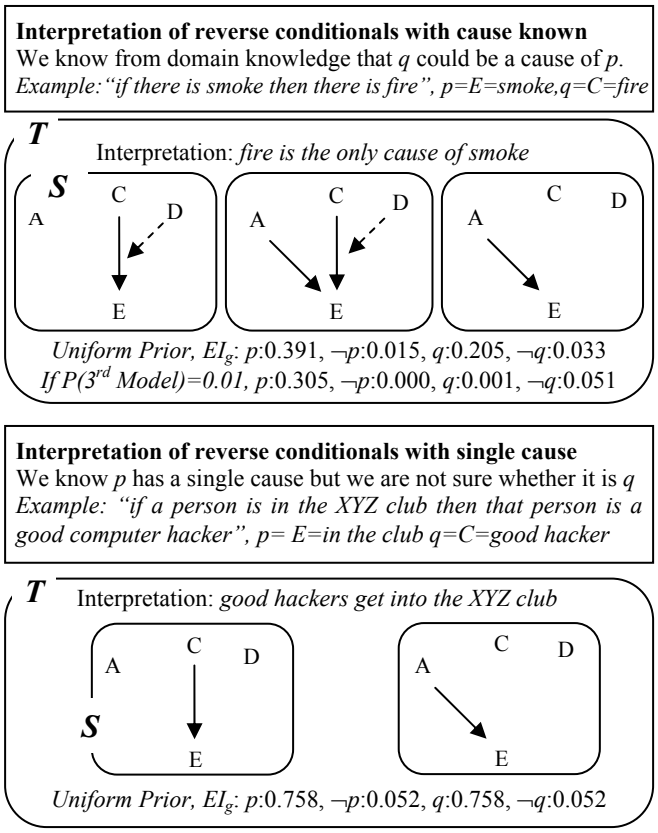


Figure 2: Predictions for reverse conditionals
Dotted arrows indicate a mixture of two hypotheses, one in which the arrow is present, and one in which the arrow is absent
All $EI_g$ values assume $P(C)=P(A)=P(D)=0.1$

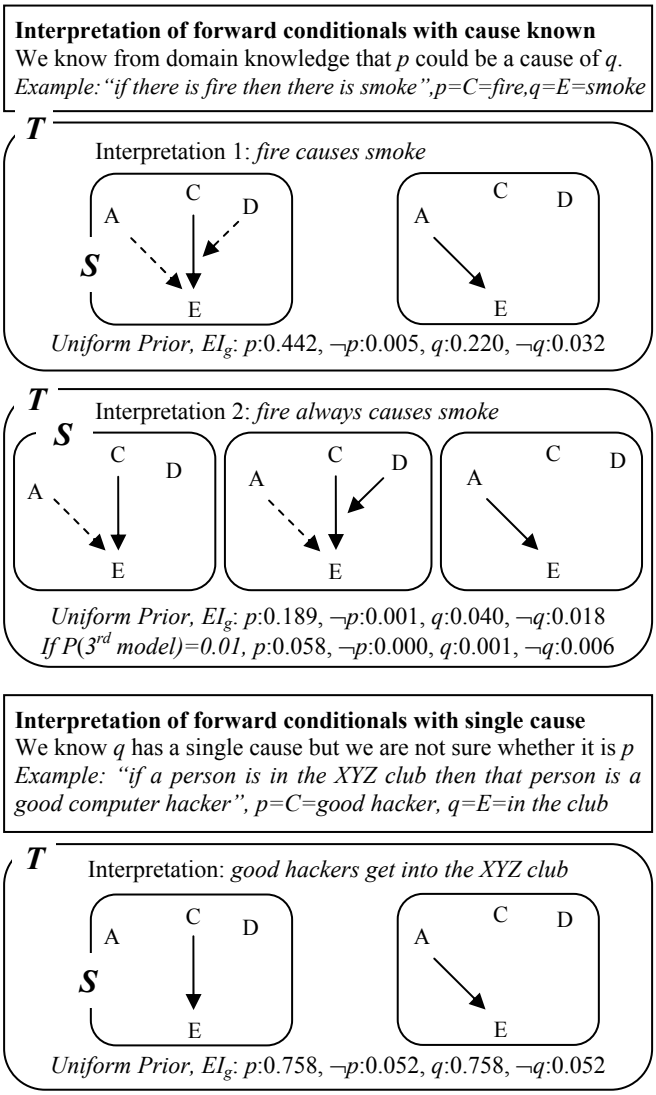


Figure 3: Predictions for forward conditionals.
Dotted arrows indicate a mixture of two hypotheses, one in which the arrow is present, and one in which the arrow is absent.
All $EI_g$ values assume $P(C)=P(A)=P(D)=0.1$

## Relation to previous analyses and phenomena

In this section we discuss how our approach accounts for previous phenomena on the selection task. We group these phenomena within a discussion of previous approaches, while highlighting the distinctive aspects of our approach.

## Information-gain (Oaksford & Chater, 1994)

In many of their publications, O&C analyze a simple comparison in which $H$ asserts complete dependency and $\neg H$ asserts a complete independency between $p$ and $q$. In our model, this is identical to the assertion that $T$ corresponds to all structures in which $p=C$, $q=E$, and $D$ does not exist, while $S$ corresponds to the subset of those structures in which a link exists from $C$ to $E$.

With the richer hypothesis space of causal models, the information gain framework predicts some previous results on the selection task that are not predicted by O&C (see next section). O&C predict that the $p$ and $q$ cards should be chosen when both $p$ and $q$ are rare and that the $p$ and $\neg q$ cards should be chosen when $p$ or $q$ is common. This is predicated on the assumptions that $I_g(p,\neg q)$ is 1, $I_g(p,q)$ is high for rare $p$ and $q$, and $I_g(\neg p,q)$ and $I_g(\neg p, \neg q)$ are zero. Our analysis suggests that these assumptions are only valid if structures in $T$ with links from $C$ to $E$ have sufficient priors. If these structures have very low priors, the $\neg q$ card should be more informative than the $q$ card because $I_{KL}(S|p,q)$ and $I_{KL}(S|\neg p,q)$ will be low, hence $EI_g(S,q)$ will be low. For example, suppose one asserts that "if a clang is heard then a pot was dropped". It is reasonable to assume that dropped pots cause clangs, hence a low prior should be placed on any structure with no link from dropped pots to clangs. Thus, finding a dropped pot that clanged ($p,q$) will not be very informative, despite $p$ and $q$ being rare, but finding a dropped fork that produced a clang ($p,\neg q$) will be very informative, hence $p$ and $\neg q$ should be chosen.

Almor & Sloman (1996) provide evidence that appears to contradict O&C (1994), in which $p$ and $q$ are rare, yet people choose the $p$ and $\neg q$ cards. Some examples of their conditionals are "*if a product gets a prestigious prize then it must have a distinctive quality*", and "*if a product breaks then it must have been used under abnormal conditions*". O&C (1996) claim these results can be accounted for using their utility-theoretic analysis of deontic tasks. However, these statements would only be deontic if they were rules that people have to follow, but it is not apparent that this is the case. According to our analysis, the conditional is reversed ($q$ causes $p$), leading subjects to interpret the statement as asserting the absence of alternate causes ($A$) while taking for granted the link from $C$ to $E$, thus assigning low probability to structures without this link. For example, since there are other possible causes of a product breaking, subjects choose the $\neg q$ card (no abnormal usage) to see if $p$ occurred (the product broke for some other reason), but there is no need to see if abnormal usage causes breakage.

A further point of differentiation is that our causal framework predicts the $\neg p$ card should be chosen in certain cases (when $I_{KL}(\neg p,q)$ and $P(q|\neg p)$ are both high).

## Social Contracts and Precautions (Cosmides, 1989)

People have been found to provide high levels of logically correct responses to Wason selection tasks about social contracts (Cosmides, 1989; Fiddick, Cosmides, & Tooby, 2000). A debate has emerged over whether this is evidence for a specialized social reasoning engine. Social contracts do indeed seem to be special, but do people reason about them differently than other tasks? In our framework, what makes them special is that they all have a consistent causal structure, in which people follow rules to ensure that $C$ produces $E$ reliably. For example, in the social contract "if you pay $10 then you get a watch", the rules compel the seller to give the buyer a watch ($E$) once $10 is paid ($C$). When the link from $C$ to $E$ is assumed to exist, as is the case in most social contract tasks, one should assign a prior of zero to any hypothesis in $T$ with this link missing. Since all the hypotheses with non-zero prior then contain links from $C$ to $E$, our causal analysis predicts that only $I_{KL}(p,\neg q)$ is high, and hence only the $p$ and $\neg q$ cards should be selected.

Precaution tasks (Fiddick et. al, 2000) have essentially the same structure as social contracts: they assume that the precaution is in force (i.e., the link from $C$ to $E$ exists), and ask subjects to determine whether the rule is being followed by everyone. Our analysis suggests that if instead the rule itself is questioned (i.e., is the rule in force?), people will interpret $S$ as asserting that there is a link from $C$ to $E$; since this link questioned, one should assign a non-zero prior to the structures in $T$ in which this link is missing, making the $q$ card useful (if $p$ and $q$ are rare) because $I_{KL}(p,q)$ is high.

The results of Fiddick et. al (2000) show that this is exactly what people do. Fiddick et. al (2000) published precaution experiments that show people choose $q$ more than $\neg q$ in "standard" versions of precaution studies such as "*if you go hunting then you wear [orange] jackets to avoid being shot*". In the "standard" version, subjects are instructed to see if it is true that the jackets are for hunting, whereas in the "precaution" version they are instructed to see if any people are endangering themselves. This result confirms that when testing whether a social contract or precaution is in force, people will test the assertion that a link exists from $C$ to $E$, and hence will choose the $p$ and $q$ cards (provided $p$ and $q$ are rare).

O&C (1994) propose a utility-theoretic account for how people make choices in social reasoning tasks. This is appropriate for tasks in which the participant is told that catching rule violators is important (i.e., has high utility). If, however, the participant is being asked simply to determine whether or not the rule is being violated, the assignment of utility to this information is not warranted. We avoid the difficulty of assigning utilities to information by using expected information gain as the sole basis on which to select cards. A causal analysis predicts the selection of $p$ and $\neg q$ responses for any task in which structures without $C$ to $E$ links are given low priors, which should be the case in all social reasoning tasks that assume the rule is in force and ask subjects to detect violators.

## Perspective Shifting

Perspective shifts (interpreting "if $p$ then $q$" as "if $q$ then $p$") have been explained as the result of adopting different perspectives on a rule – the enforcer vs. actor. We propose that perspective shifts occur when three conditions are met: (1) $C$ is a known cause of $E$, (2) it is not obvious whether $D$ exists, and (3) it is not obvious whether $A$ exists. This sets

up the hypotheses in Figure 4. For example, "if you pay $10 then you get a watch" can be shifted to "if you got a watch then you paid $10" because our domain knowledge tells us that no disabling conditions exist (the buyer must get the watch once $10 are paid), and no alternate causes exist (the buyer cannot get the watch without paying).

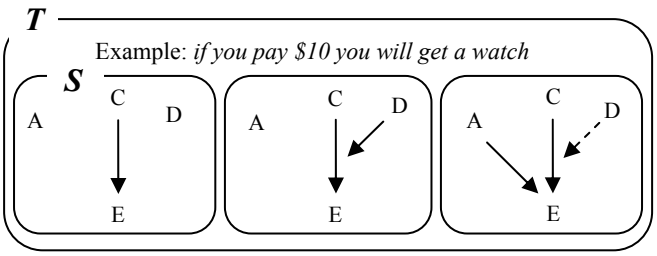


Figure 4: Perspective shifting hypotheses

On this view, perspective shifting occurs not just on deontic tasks, but in any situation in which the above three conditions are met. For example, "if water boils then it is over 100°C" could easily be interpreted to imply that "if water is over 100°C then it will boil". Perspective shifting can therefore occur, even in non-deontic situations, when the asserted structure does not contain *D* or *A* links.

### Necessity and Sufficiency (Ahn & Graham, 1999)

Ahn & Graham (1999) show that most people choose the normative response if it is clear that the statement asserts either that *p* is a sufficient condition for *q* or that *q* is a necessary condition for *p*, or both. Asserting that *p* is a sufficient condition for *q* corresponds to asserting that *D* does not exist. For example, asserting that *flipping the switch* is sufficient for the *lights turning on* corresponds to asserting that nothing (*D*) can disable the switch. In contrast, asserting that *q* is a necessary condition for *p* corresponds to asserting that *A* does not exist in the causal model. For example, asserting that *flipping the switch* is necessary for *lights turning on* corresponds to asserting that nothing else (*A*) could turn on the lights. Both of these cases assume that the link from *C* to *E* exists, hence as before, *p* and $\neg q$ are the most informative cards (or *q* and $\neg p$ when a conditional with "must" is reversed to say "may"), which follows Ahn & Graham's predictions. Ahn & Graham (1999) also discuss cases in which *p* is asserted to be both necessary and sufficient for *q*, in which cases subjects choose all 4 cards. This corresponds to asserting that neither *A* nor *D* exist, and $I_{KL}(S|\neg p,q)$, $I_{KL}(S|p,\neg q)$ are both high.

An open question in Ahn & Graham's (1999) theory is how people know whether *p* is necessary or sufficient for *q* in cases when it is not explicitly stated. A cause can be necessary or sufficient for an effect, but it does not make sense to say that an effect is necessary or sufficient for a cause (e.g., a clang could not be necessary or sufficient for a pot to drop, because dropped pots precede clangs). Because of this causal asymmetry, some amount of causal reasoning must precede determination of necessity and sufficiency relationships. Furthermore, determining necessity or sufficiency can be done using just causal knowledge, as *p* is necessary for *q* if *A* does not exist, and *p* is sufficient for *q* if *D* does not exist.

## Conclusion

Causal reasoning underlies many of our intuitive judgments in everyday life, and the results we present here demonstrate that causal structure plays an important role in a domain of reasoning previously thought to be governed by logic and probability. Our approach predicts a number of effects on the selection task that do not follow naturally from previous approaches. If used appropriately, the selection task is an excellent tool for testing people's abilities to gather evidence and become more informed about their world. Since knowing the causal structure of the world is of great value for making predictions in every life, it is perhaps not surprising that the cards people naturally select tend to be those that maximize the amount of knowledge that can be obtained about causal structure from a single observation.

## References


Ahn, W., & Graham, L. M. (1999). The impact of necessity and sufficiency on information choices in the Wason four-card selection task. *Psychological Science*, 10, 237-242

Ahn, W. (1999). Effect of Causal Structure on Category Construction. *Memory & Cognition*, 27, 1008-1023

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation vs. mechanism information in causal attribution. *Cognition*, 54, 299-352.

Almor, A. & Sloman, S. A. (1996). Is deontic reasoning special? *Psychological Review*, 103(2): 374-380

Cosmides, L. 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition,* 31, 187-276

Fiddick, L., Cosmides, L., Tooby, J. (2000). No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task. *Cognition*, 77, 1-79

Glymour, C. & Cheng, P. (1998). Causal mechanism and probability: a normative approach. In *Rational models of cognition,* Oaksford, M. & Chater, N., Oxford University Press.

Krynski, T.R., Tenenbaum, J. B.. (2003). The Role of Causal Models in Reasoning Under Uncertainty. *Proceedings of the 25th Annual Conference of the Cognitive Science Society.*

Oaksford, M. and Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review* 101, 608-631

Oaksford, M. & Chater, N. (1996). Rational Explanation of the Selection Task. *Psychological Review*, 103 (2), 381-391

Over, D. & Jessop, A. (1998). Rational analysis of causal conditionals and the selection task. In *Rational models of cognition,* Oaksford, M. & Chater, N., Oxford University Press.

Pearl, J. (2000). Causality: models, reasoning, and inference. Cambridge University Press

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science,* 27, 709-748

Wason, P.C. (1996). Reasoning. In B.M. Foss (Ed.), *New Horizons in Psychology*, (pp. 135-151). Harmondsworth, Middlesex, England: Penguin.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 53-76.



**Acknowledgements** We thank Brigid Dwyer, Sarah Newton, and Suzanne Luther for their enthusiastic assistance. JBT was supported by the Paul E. Newton chair. TRK was supported by a graduate fellowship from the National Science Foundation.