Inferring non-observed correlations from causal scenarios: The role of causal knowledge

José C. Perales, Andrés Catena,* and Antonio Maldonado

Departmento de Psicología Experimental, Facultad de Psicología, Campus de Cartuja s/n,
University of Granada, 18071 Granada, Spain

Received 1 April 2003; received in revised form 10 July 2003

Abstract

This work aimed at demonstrating, first, that naïve reasoners are able to infer the existence of a relationship between two events that have never been presented together and, second, the sensitivity of such inference to the causal structure of the task. In all experiments, naive participants judged the strength of the causal link between a cue A and an outcome O in a first phase and between a second cue B and the same outcome O in a second phase. In the final test, participants estimated the degree of correlation between the two cues, A and B. Participants perceived the two cues as significantly more highly correlated when they were effects of a common potential cause (Experiment 1a and 2) than when they were potential causes of a common effect (Experiment 1b and 2). This effect of causal directionality on inferred correlation points out the influence of mental models on human causal detection and learning, as proposed by recent theoretical models.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Human learning; Causal learning; Mediated learning; Contingency; Causality

Causal relationships learning is one of the most basic psychological abilities. Adaptation to the environment requires being able to predict the consequences of the

* The research described here was supported by a grant from the Ministerio de Educación, Cultura y Deportes (Programa de Formación de Personal Investigador y Profesorado Universitario), Spain, for José C. Perales, and by a DGICYT grant for Andrés Catena and Antonio Maldonado. We are also grateful to Antonio Cándido, David Shanks, Patricia Cheng, and two anonymous reviewers for their many helpful comments.

*Corresponding author. Fax: +34-958-246239.
E-mail address: acatena@ugr.es (A. Catena).
observed causes and to identify the causes of the observed outcomes; that is, behaving in accordance with the “causal texture” of the environment (Tolman & Brunswick, 1935).

Causal relationships are not evident, given that the mechanisms by which causes and effects are connected remain hidden to our senses. However, two events that are causally linked tend to covary: if event A causes an outcome O, A, and O will frequently appear together. In order to learn a causal relationship between two events, it is necessary to determine whether the observed covariation is due or not to a causal relationship between these events, because every observed covariation in the world is not causal. For instance, the gauge of a barometer predicts rain, as far as both variables are correlated, because they have a common cause. However, the barometer gauge does not cause the rain. Covariation (also referred to as correlation or statistical relevance; See Cheng, 1993) can be regarded as the main sensory input for causal induction processes (Cheng, 1997). Nevertheless, correlation itself is not sufficient to infer a causal link. Further processes are needed to discriminate between true (causal) and spurious (non-causal) correlation.

The ability to discriminate between causal and non-causal correlations requires complex learning and conceptual change processes, developing across consecutive stages of human development. Those high-level processes involve previously learned causal knowledge and the application of normative principles. They operate over learned correlations to decide their causal nature and determine how contingency is computed.

Most authors in the field of causal learning assume that basic learning mechanisms, common to a variety of learning situations, are partially responsible for covariation and causal learning. In fact, in a variety of causal learning experiments, the task structure closely resembles that of conditioning tasks. In the standard procedure, the task is divided into several discrete temporal intervals (trials), in which two events can occur. The first event is usually called cue or predictor, and the second one outcome. From the combination of the presence and absence of the cue and the outcome, four types of trials can result: in a type trials, both the cue and the outcome are present; in b type trials, the cue is present, but the outcome is absent, in c type trials, the cue is absent, but the outcome is present; and in d type trials, both the cue and the outcome are absent. The frequencies of these four types of trial are used to compute the degree of relationship between the cue and the outcome. A common measure of between-events relationship (contingency) is \( \Delta P \), an estimate of the difference between the probability of the outcome when the cue is present and the probability of the outcome when the cue is absent. Therefore, \( \Delta P \) indicates the degree to which the cue signals an increment in the probability of the outcome:

\[
\Delta P = \frac{a}{a + b} - \frac{c}{c + d},
\]

where \( a, b, c, \) and \( d \) stand for the absolute frequencies of the four types of trial.

---

1 According to Piaget, normative manipulation principles (by virtue of which, for example, extraneous factors are controlled while manipulating the target factor) are not acquired until the emergence of operational intelligence (Piaget, 1936).
The objective contingency level influences the perceived contingency between the cue and the outcome (see, for example Alloy & Abramson, 1979; Chatlosh, Neunaber, & Wasserman, 1985; Jenkins & Ward, 1965; Wasserman, 1990; Wasserman, Elek, Chatlosh, & Baker, 1993), but also the perceived causal relationship between them; and similar results are obtained in causal and contingency learning tasks.²

The similarity between results from causal and contingency judgment tasks supports the idea that the two are comparable. Several theories hold that contingency learning depends on mechanisms that are independent of subjects’ previous knowledge (i.e., associative mechanisms), maintaining that learning depends on the temporal and physical features of the task. However, a second type of theory assumes that learning mechanisms are constrained by previous causal knowledge (i.e., the Causal-Model theory). Henceforth, we will refer to the former type as data-driven theories, and to the latter one as cognitively-driven theories.

Associative theories describe covariation learning as a conditioning-like process, in which the successive cue-outcome pairings progressively strengthen the connection between the mental representations of the cue and the outcome (See Allan, 1993, for reviews of data supporting this approach; Price & Yates, 1995). Although associative learning models incorporate the influence of several attentional and strategic factors on learning (See Shanks, 1996, for a review; see also Mackintosh, 1975; Pearce, 1987; Pearce & Hall, 1980; Van-Hamme & Wasserman, 1994), the learning process is considered automatic and data-driven (Cobos, Caño, López, Luque, & Almaraz, 2000; Matute, Arcediano, & Miller, 1996; Shanks & López, 1996).

According to the Causal-Model approach, previous causal knowledge influences learning mechanisms to the point that between events causal relationships are learned directionally,³ in such a way that causal inference incorporates the intrinsic asymmetry of causal relations (Waldmann, 2000, 2001; Waldmann & Holyoak, 1992, 1997). Support for this proposal comes from the interaction between causal directionality (predictive versus diagnostic), manipulated by means of instructions, and cue competition effects.

In the last three decades, cue competition effects have been considered a key tool to discriminate between different learning theories. In general terms, both in animal and human learning paradigms, it has been proved that the perceived covariation of a cue and an outcome depends, not only on the contingency between both events, but also on the contingency between other cues and the same outcome, as shown...
by the well-known blocking effect (Chapman & Robbins, 1990; De Houwer, Beckers, & Glaui, 2002; Kamin, 1969; Lovibond, Siddle, & Bond, 1988; Shanks, 1985).

However, from a normative point of view, blocking is considered ‘rational’ only if cues are defined as potential causes, and outcomes as effects (the so-called predictive tasks; see Waldmann & Holyoak, 1992, 1997, for detailed discussions on this issue). The Causal-model account maintains that human causal learning is sensitive to these types of normative principles, and, thus, it predicts blocking in those situations in which the competitive cues are defined as potential causes (predictive tasks), but not when they are defined as effects (diagnostic tasks).

Unfortunately, whether cue competition is sensitive or not to causal directionality has become a controversial issue, lasting for more than a decade (Cobos et al., 2000; Matute et al., 1996; Shanks & López, 1996; Waldmann, 2000, 2001; Waldmann & Holyoak, 1992, 1997).

In this work, a different methodological perspective has been adopted. Learning from direct experience is not the only way to acquire knowledge about correlation between cues. As shown later, during or after learning a set of relationships, new relationships can be derived from this knowledge. For example, in several experiments with rats, Holland (1981, 1998) demonstrated that after limited exposure to the relationship between an auditory CS and food, pairings of the same CS with the toxin LiCl, in the absence of food, were sufficient to establish an aversion to the food US that had been previously paired with that CS. This effect has been called representation-mediated conditioning. The author provides an associative explanation of this effect: An associative link between CS and the appetitive US is strengthened during the first learning task. In the second one, the CS activates the representation of the appetitive US simultaneously with the onset of the aversive US (LiCl). The joint occurrence of the aversive US and the representation of the appetitive US establishes an associative link between both USs. In other words, during the second phase the appetitive US becomes a CS for the aversive US.

In human causal learning tasks, the situation is rather more complex, as temporal and causal order of the events presented during the task can be decoupled. Therefore, this type of mediated learning can be used to clarify the role of causal knowledge in contingency learning.

The main aim of the reported experiments was, first, to ascertain whether people inferred the existence of correlation between two cues that have never been directly paired during the task, and, second, whether or not the inference of such correlation is sensitive to causal directionality. To this end, in the first phase, a cue (the appearance of a chemical substance in the blood flow of a series of fictitious patients) was paired with an outcome (a fictitious neurological disease). In the second phase, a different cue (another chemical substance) was also paired with the same outcome (the disease). In each phase, participants were asked to estimate the degree to which the cue and the outcome were causally related. At the end of the experiment, participants were unexpectedly asked to estimate the strength of the relationship between the two cues.

Cues and outcome can be both causes and effects, depending on the causal directionality condition. In the predictive condition, the cues (the two chemical
substances) were defined as potential causes of the outcome (the disease). Conversely, in the diagnostic condition, cues were defined as effects of the disease.

There are two ways by which this non-observed correlation can be inferred. First, it could be computed during the second learning phase, by means of an associative mechanism, in a similar way to Holland's (1981, 1998) proposal to account for mediated conditioning. It is important to note that associative mechanisms are not affected by causal knowledge, i.e. causal directionality. Therefore, similar inferred correlations in both predictive and diagnostic tasks are expected.

Second, non-observed correlations could be inferred from a mental model built during or after learning. We assume that a mental model is a working memory representation of real (or imaginary) situations, as a result of perception and/or comprehension of instructions (Johnson-Laird, 1983; Marr, 1982). Given that their structure is isomorphic to the situation, mental models can be used to make inferences. There are a great number of possible causal mental models, the common-cause and the common-effect being the most important in our experimental preparation (see Fig. 1).

If naïve reasoners are able to build models like these, by integrating the information provided by instructions (causal directionality) and the information provided during the task (chemical 1—Outcome and chemical 2—Outcome objective contingencies), correlation between the two cues can be deduced in a normative manner. In the predictive condition, the same effect \( (E) \) is produced by two different causes (\( C_1 \) and \( C_2 \)), and, therefore, the common-effect model applies. For example, lung cancer (\( E \)) can be produced either by smoking (\( C_1 \)) or by contamination from cars (\( C_2 \)). There is no way to compute the degree of correlation between the two causes. In fact, the joint distribution function, \( p(C_1, C_2, E) \) does not exist in our procedure, because,

\[
p(C_1, C_2, E) = p(E/C_1, C_2)p(C_1)p(C_2),
\]

and \( p(E/C_1, C_2) = 0 \), given that \( C_1 \) and \( C_2 \) are never presented together. Of course, a model that makes it possible to answer such question could be built. For example, it can be assumed that causes are independent (noisy-or schema, see Appendix A), and that its degree of correlation is 0, regardless of the programmed contingency between each cause and the effect.

Because in the diagnostic condition two different effects (\( E_1 \) and \( E_2 \)) are caused by the same cause (\( C \)), the common-cause model applies. In this case, the joint

---

Fig. 1. Two possible causal models underlying the predictive (B) and the diagnostic (A) task in Experiments 1a, 1b, and 2. A and B (between brackets) are the cues, and O, the outcome. Letters C and E represent the causal roles of the stimuli as potential causes and effects.
distribution function is defined as \( p(C, E_1, E_2) = p(E_1/C)p(E_2/C)p(C) \). Then, it follows that the degree of correlation between the two effects depends on the correlation of each one with the cause. A positive degree of correlation is expected whenever both programmed contingencies (chemical 1—Outcome and chemical 2—Outcome) are higher than 0 (see Pearl, 2000; Waldmann & Martignon, 1998). For example, a throat bacterial infection produces both throat pain and fever, and, hence, a high correlation between these two symptoms is expected.

There exist, at least, two other types of models that can be applied to some of the situations presented in our experiments. These models are the causal-chain model, in which the three stimuli form a causal sequence (for example, one cue is the cause of the other cue, and the second cue is the cause of the outcome), and the interactive model, in which the two causes interact to generate or prevent the appearance of the effect. The possible application of these two models cannot be discarded a priori (except in the case in which they are logically incompatible with the programmed contingencies). The observed pattern of data in the three experiments will show that naïve subjects rarely apply such models.

In summary, whereas associative models predict inferred correlation to be insensitive to causal directionality, the most plausible version of a causal-model theory predicts that reasoners should infer the existence of correlation between the cues only in the diagnostic condition.

**Experiment 1a**

Our first experiment was designed to study whether people derive a correlation between two cues that have never been paired, when the cues are defined as possible effects of a common cause in a diagnostic causal learning task. Participants were instructed to imagine being members of a scientific team interested in evaluating early symptoms of a new neurological disease. Two predictors or cues—two chemical substances either present or absent in the blood flow of a series of fictitious patients—were separately paired with the same outcome (the disease), in two learning phases. Contingencies between each cue and the outcome were manipulated in order to test their potential effect on inferred correlations. After the two training phases, subjects were unexpectedly asked to estimate the correlation between the two chemical substances.

**Method**

**Participants and apparatus**

The participants were 126 University of Granada students. Eighty six of them were female and their ages varied between 18 and 26, with a median of 21. They obtained course credits for their participation. Participants were tested individually and required between 20 and 30 min to complete the experiment. The instructions, stimuli and scale judgments were displayed on a PC computer screen. Participants responded by pressing the computer keyboard or mouse buttons.
Design and procedure

After the participants had typed in their age and gender, they were presented with the instructions. The instructions biased the participants to consider two cues (two substances, XDA and RVO, that may or may not be in the blood flow of each patient of a series of 32) as possible effects of an outcome (a fictitious disease called Riskendt’s syndrome).

The experiment was divided into two phases (see Table 1). In phase 1, participants were exposed to some degree of contingency between the first chemical (XDA for half of the subjects and RVO for the other half) and the syndrome. In phase 2, the same group of subjects was exposed to some degree of contingency between the second chemical (RVO for subjects that received XDA in the first phase, and XDA for the rest of the subjects) and the same disease. In the final test phase, the participants were required to estimate the strength of the correlation between the two chemical substances. For technical reasons, three participants were unable to give this judgment (two in group 3 and 1 in group 9, see Table 1). The question (translated from Spanish) was worded as follows:

If the presence of the two chemicals were tested simultaneously in a new sample, to what degree would they appear together? (−10, they would never appear together; +10, they would always appear together. Intermediate numbers indicate intermediate levels of correlation).

In both phases, the series of patients were presented in a trial-by-trial mode. Each record indicated whether the target chemical substance was detected in the blood flow of that patient or not and, after 1500 ms, whether that patient was diagnosed with the Riskendt’s syndrome or not. To go on to the next trial, participants clicked a button on the screen labelled “Next Trial” or pressed Enter.

Table 1
Design and between cues mean correlation judgments of Experiments 1a and 1b

<table>
<thead>
<tr>
<th>Condition</th>
<th>Learning phase 1</th>
<th>Learning phase 2</th>
<th>Judgment A–B</th>
<th>Judgment A–B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A→Riskendt</td>
<td>B→Riskendt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic task</td>
<td>Predictive task</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>−.50 (3.87)</td>
<td>−1.40 (3.37)</td>
</tr>
<tr>
<td>2</td>
<td>.50</td>
<td>1.92 (3.87)</td>
<td>−1.08 (4.33)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>.78 (4.85)</td>
<td>1.04 (3.06)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.50</td>
<td>.00</td>
<td>−1.29 (6.79)</td>
<td>.25 (5.29)</td>
</tr>
<tr>
<td>5</td>
<td>.50</td>
<td>1.47 (2.55)</td>
<td>−.45 (4.95)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
<td>3.26 (2.63)</td>
<td>1.81 (3.14)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.0</td>
<td>.00</td>
<td>−.12 (4.30)</td>
<td>−1.54 (3.88)</td>
</tr>
<tr>
<td>8</td>
<td>.50</td>
<td>1.96 (2.16)</td>
<td>−1.37 (5.45)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.0</td>
<td>6.73 (3.71)</td>
<td>1.41 (6.55)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Standard deviations are between brackets. A and B are fictitious chemical substances defined as possible symptoms (effects) of the same cause (the Riskendt’s syndrome) in Experiment 1a, and as possible causes of the same effect (the Riskendt’s syndrome) in Experiment 1b. Judgments were always estimated in a scale ranging from −10 to +10.
After every eight-trial block, participants were asked to estimate the degree to which the target chemical was caused by the disease. Participants made their estimation by using a horizontal 20 cm scrollbar scale with a pointer initially located at its center. The scale pointer could be moved in both directions (within a range from −10 to +10) by clicking on the scrollbar or using the cursor keys in the keyboard. The second phase was procedurally identical to the first one, except for the level of contingency between the chemical and the disease, which was established by a factorial design.

A between-subjects design, in which each participant was randomly assigned to one of 9 experimental conditions (N = 14), was used. In the first phase, contingency between a fictitious chemical substance (XDA or RVO, henceforth substance A) and the fictitious disease (Riskendt’s syndrome) was manipulated across three levels of contingency, \( \Delta P = 0, .5 \) or 1, and, in the second phase, contingency between the other chemical (RVO or XDA, henceforth B) and the same outcome was also manipulated across the same three levels. These variables were combined according to a factorial design. The null contingency level (\( \Delta P = .00 \)) consisted of 8 trials of each type of trial (a, b, c, and d); the .5 contingency level was composed of 12 type a (cue-outcome) trials, 4 type b (cue-no outcome) trials, 4 type c (no cue-outcome) trials and 12 type d (no cue–no outcome) trials; and finally, the high-contingency level (\( \Delta P = 1.00 \)) was composed of 16 type a trials and 16 type d trials.

Results and discussion

Causal judgments for A and B

All statistical decisions in this and the subsequent experiments were made using an alpha level of .05. Two independent variables were considered for the present analysis. Judgment order refers to the order in which judgments were made. Each subject made a causal judgment at the end of each learning phase, which means that this variable was manipulated across two levels: first and second judgment. The other manipulated variable was cue-outcome contingency, measured as \( \Delta P \). A 2 × 3 (Order × Contingency) mixed analysis of variance (ANOVA) applied to the last causal judgment in each phase of the task revealed a significant effect of Contingency, \( F(2,242) = 50.70 \) (\( M = 7, 27, \) and 69, for contingencies 0, .50, and 1.00, respectively). Neither the effect of judgment order nor the interaction Order × Contingency was significant (all \( F < 1 \)). LSD post-hoc tests of Contingency showed significant differences between .00 and .50 levels, as well as between the .50 and 1.00 levels.

These results show that causal estimates depended on the objective contingency between cue and outcome (for the present purposes, both phases can be considered as independent causal learning tasks), as there is no influence from the first to the second phase. They also suggest that the procedure was adequate to establish a causal link between a potential cause and an effect.

Inferred correlation

Between-cues (between-effects) judgments were analysed by applying a 3 × 3 (A-Outcome contingency × B-Outcome contingency) between-subjects ANOVA. The analysis showed a significant effect of B-Outcome contingency, \( F(2,114) = 10.88 \).
The effects for A–Outcome contingency, $F(2,114) = 3.02, p = .053$, and the interaction between the two factors, $F(4,114) = 2.26, p = .066$, were marginally significant. However, the Levene’s test showed that error variances differed among groups, $F(8,114) = 4.21$. Therefore, we also carried out Kruskal–Wallis’s non-parametric tests, establishing $p$ by the Bonferroni’s error correction method. Although this strategy is highly conservative, it most significant clearer results than those of the ANOVA.

In all Kruskal–Wallis analyses, an alpha level of .008 was established. The effect of A–O contingency manipulation in each level of B–O contingency, and the effect of B–O contingency in each A–O contingency level were analysed. Only the effect of A–O contingency in the highest B–O contingency level, and the effect of B–O contingency in the highest A–O contingency level were significant, $H(2) = 11.56$, and $H(2) = 15.42$, respectively. Detailed analyses showed significant differences between conditions 3 (A–O $\Delta P = 0$ and B–O $\Delta P = 1$) and 9 (A–O $\Delta P = 1$ and B–O $\Delta P = 1$), and between condition 9 and conditions 7 (A–O $\Delta P = 1$, B–O $\Delta P = 0$) and 8 (A–O $\Delta P = .5$, B–O $\Delta P = .5$).

In general, the present results suggest that, when one of the contingencies is held constant at a highly positive value, the perceived contingency between the two effects tends to be a function of the value of the other contingency; but whenever one of them has a null or low value, there is no perception of correlation between both effects. This pattern of results is consistent with the idea that the combination of individually learned causal links involving a common cause can induce people to believe that the two effects of that cause are correlated to some degree. A stronger test of this idea will be found in Experiment 2.

**Experiment 1b**

Our second experiment was procedurally identical to the first one. The same cues and the same outcome were used, and, as in the previous experiment, in each trial the presentation of the chemical substance preceded the presentation of the syndrome. Therefore, all the temporal and physical features of the tasks were identical to those in the first experiment. However, in this case, participants were instructed to consider the chemical substances as potential causes of the disease. Thus, causal order matched temporal order. As described before, in this situation the participant is expected to infer that the effect has two possible causes. From that model, there is no way to normatively infer the existence of correlation between the two chemicals as far as they are perceived as independent causes. However, if the relationship between the two cues is formed associatively, there is no reason to expect the results of this experiment to differ from those obtained in the previous one.

**Method**

**Apparatus and participants**

The participants were 108 University of Granada students. Seventeen were male and 91 were female, and their ages varied between 18 and 36, with a median of 21.
All the other apparatus and participants details were the same as in the previous experiment (except for the number of subjects in each group, \( N = 12 \)).

**Procedure, design, and variables**

The procedure and design, as well as the variables manipulated and registered, were identical to those described previously. The occurrence of each chemical was described in the instructions as a potential cause of the disease. After every eight-trial block, participants were asked to estimate the degree to which the chemical caused the disease. The order in which the cues and the outcome were shown to the participants in each trial was the same described in Experiment 1a.

**Results and discussion**

*Causal judgments for A and B*

A 2 x 3 two-way (Order x Contingency) analysis of variance (ANOVA) on the last judgment of each phase revealed a significant effect of Contingency (\( M = 5, 30, \) and 53, for \( .00, .50, \) and \( 1.00 \) contingencies, respectively), \( F(2, 204) = 22.49 \). Neither the effect of Cue Order nor the interaction Cue of order x Contingency was significant (\( F < 1 \) in both cases). Contingency main effect’s trend analysis showed that only the linear component was significant, \( F(1, 204) = 44.90 \). Post hoc LSD comparisons showed significant differences between \( .00 \) and \( .50 \) contingency levels, as well as between \( .50 \) and \( 1.00 \) contingency levels.

These data demonstrate that the procedure is adequate to establish strong beliefs about the causal relationships between each cue and the outcome. However, causal judgments in this experiment were slightly lower than in the previous one. We will discuss the potential implications of this apparent difference later.

*Inferred correlation*

As in the previous experiment, two independent variables were considered for the present analysis: A-Outcome contingency and B-Outcome contingency, measured by \( \Delta P \). For technical reasons, three participants failed to give a judgment (two of group 1 and 1 of group 3, see Table 1). A 3 x 3 between-subjects analysis of variance (ANOVA) of the judgments of correlation between the two chemicals at the end of the experiment did not reveal any effect of A-Outcome contingency (\( F < 1 \)) or the interaction between the two factors (\( F < 1 \)), although the effect of B-Outcome contingency was not far from being significant, \( F(2, 96) = 2.67, p = .074 \).

As can be seen in Table 1, inferred correlations in this experiment were much lower than in the previous one, and the interactive pattern observed in the previous experiments disappeared completely.

**Discussion of experiments 1a and 1b**

The pattern of results of experiments 1a and 1b suggests that causal directionality plays an important role in determining the degree to which naive reasoners infer correlation between non-paired cues from a previously learned set of causal
relationships. However, there are several reasons why we should be cautious about this interpretation.

First, the effect of causal directionality has been obtained only between experiments. Although both experiments are identical in most senses, there are some differences that are important to take into account. On the one hand, in Experiment 1a, 14 participants were included in each condition, whereas only 12 participants were included in Experiment 1b. Obviously, this cannot be argued to be the reason why the patterns from those experiments are so radically different, but it could have affected the possibility of finding significant differences in Experiment 1b. This point is important for ascertaining whether the differences between experiments are qualitative, as predicted by the all-or-nothing logic of the causal-model theory, or quantitative, as predicted by integrative models. On the other hand, causal judgments about the relationships between each cue and the outcome were slightly lower in Experiment 2. It could be argued that the lack of inferred correlations in Experiment 1b is not due to a true effect of causal directionality, but to a previous failure to establish strong causal beliefs.

Second, although the interpretation of patterns in Experiments 1a and 1b is straightforward, simple cross-experimental comparisons seem to show that the main difference between both experiments depends on global judgments (for the experiments taken as a whole) and on judgments in high-contingency conditions (for condition-to-condition cross-experimental comparisons). In other words, only when causes are deterministic ($\Delta P = 1$), is the effect of causal directionality on mediated learning statistically significant. It remains unknown whether or not this effect would appear with non-deterministic causes.

And finally, it could be argued also that the cover story about causes and effects of certain diseases could trigger the application of a very narrow mental model, generalized from subjects’ experience on that particular matter. Participants would be applying their previous knowledge about correlation between symptoms of a disease, rather than a general statistical principle. In Experiment 2, a different, less familiar, context was described in the instructions, in order to increase the generality of our findings.

**Experiment 2**

Experiment 2 was designed to overcome the limitations of Experiments 1a and 1b. A less familiar causal context was used in the cover story. In the diagnostic condition, two fictitious micro-organisms (called Ladiarium and Espiridia) were described as potential causes of the appearance of a certain chemical substance (Ascetileno) in the air of a sample of cities that were analysed for health research purposes. Conversely, in the predictive condition, the proliferation of each micro-organism was described as a potential effect of the appearance of the chemical in the sample of cities. The design was simplified in order to make the analyses of distributions easier. In this case, only two contingency conditions were compared for each causal structure.
Method

Apparatus and participants

The participants were 64 University of Granada students. Eleven were male and 55 were female, and their ages varied between 18 and 40, with a median of 22. All apparatus and participant's details were the same as in Experiments 1a and 1b.

Procedure

The task was divided into two phases, as in the previous experiments. In the first phase, one of the micro-organisms (Espiridia for half of the subjects and Lariadium for the other half, A cue) was paired with the appearance of Ascetileno in the air of a sample of cities. In the second phase, the same training procedure was used to pair the other micro-organism (Lariadium for half of the subjects, and Espiridia for the other half, B cue). Each trial started with the presentation of the cue (the micro-organism), or its absence. The presence of the cue was indicated by a schematic drawing and the message “Lariadium (or Espiridia) was detected in this city.” The absence of the cue was indicated by the absence of the drawing in the corresponding box, and the message “Lariadium (or Espiridia) was NOT detected in this city.” After 1500 ms, a drawing and a written message informed the participant about the presence or the absence of the chemical substance Ascetileno.

In the high-contingency group ($\Delta P = .77$), each phase consisted of 16 type a trials, 2 type b trials, 2 type c trials, and 16 type d trials. In each phase, the relationships between cue A and the outcome between and cue B and the outcome were high but non-deterministic. Also, half of the subjects of this group worked in a diagnostic task, and the other half worked in a predictive task. In the mixed-contingency group, one of the learning phases (the first one for a half of the subjects, and the second one for the other half) was composed of 9 trials of each type ($\Delta P = 0$), thus yielding a null contingency level. The cue-outcome contingency was .78 in the other phase (the second one for half of the subjects, and the first one for the other half). Again, half of the subjects of this group worked in a diagnostic, and the other half worked in a predictive task.

In summary, there were four groups ($N = 16$). The two high-contingency groups received high-contingency training with both cues; one of the groups worked in a predictive task, and the other worked in a diagnostic task. The two mixed-contingency groups received high-contingency training with one of the cues and null contingency training with the other; one of them worked in a predictive task, and the other worked in a diagnostic task. Therefore, two between-subjects factors were manipulated: Contingency (high vs mixed) and causal directionality (predictive vs diagnostic).

Results and discussion

Causal judgments

Two factors were considered independently: judgment order and contingency. As in previous analyses, judgment order did not yield any significant effect on causal
judgments, and was not considered for further analyses. The only significant difference was the one involving mean judgments for the high and the low-contingency phase of the experiment in the mixed-contingency group (see Table 2). \( F(1, 31) = 43.995 \).

It is important to note that the high-contingency condition judgments in the predictive group were slightly—but not significantly—higher than those in the diagnostic group (just the opposite trend, compared to previous experiments). In other words, causal links are not more poorly learned in the predictive than in the diagnostic task. Consequently, the higher between-cues correlations in the diagnostic condition (Experiment 1a) than in the predictive condition (Experiment 1b) can not be attributed to a failure to establish strong causal beliefs.

**Inferred correlations**

Preliminary analysis including stages order failed to show any reliable effect; therefore, we omit this variable in the rest of the analysis. Judgments for correlation between the cues were introduced in a \( 2 \times 2 \) (contingency condition \( \times \) causal directionality) between-subjects ANOVA. Mean judgments for each group are displayed in Table 3.

Statistical analysis showed a significant main effect of contingency manipulation, \( F(1, 60) = 26.596 \), but not of causal directionality \( (F < 1) \). The interaction between

<table>
<thead>
<tr>
<th>Mixed-contingency condition</th>
<th>Low-contingency stage (( \Delta P = 0 ))</th>
<th>High-contingency stage (( \Delta P = .78 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic</td>
<td>1.688</td>
<td>6.375</td>
</tr>
<tr>
<td>Predictive</td>
<td>1.313</td>
<td>6.750</td>
</tr>
<tr>
<td>High-contingency condition</td>
<td>High-contingency stage 1 (( \Delta P = .78 ))</td>
<td>High-contingency stage 2 (( \Delta P = .78 ))</td>
</tr>
<tr>
<td>Causal direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic</td>
<td>6.563</td>
<td>6.438</td>
</tr>
<tr>
<td>Predictive</td>
<td>6.875</td>
<td>6.813</td>
</tr>
</tbody>
</table>

*Note.* Stages in the mixed-contingency condition have been defined according to the contingency level. Stages in the high-contingency condition have been defined according to their temporal order.

<table>
<thead>
<tr>
<th>Causal directionality</th>
<th>Contingency condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mixed (.00/.78)</td>
<td>High (.78/.78)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>.750 (2.46)</td>
<td>0</td>
<td>5.688 (1.81)</td>
</tr>
<tr>
<td>Predictive</td>
<td>1.438 (2.13)</td>
<td>0</td>
<td>3.438 (3.88)</td>
</tr>
</tbody>
</table>
the two factors was significant, $F(1, 60) = 4.768$. Post hoc LSD tests proved that all group-to-group simple comparisons were significant, except the one between the two mixed-contingency groups (diagnostic vs predictive). No effects of stages order in the mixed-contingency groups were observed.

Therefore, an effect of causal directionality on inferred correlation was found. However, this effect was a quantitative rather than a qualitative one. In other words, in the high-contingency condition ($\Delta P = .78$), in which learned causal links are strong, reasoners are more likely to infer that the two cues are correlated if the cues are effects of a common cause than if they are causes of a common effect. However, mean inferred correlation for the predictive condition (in which the two cues are causes) is also higher than the baseline defined by the mixed-contingency groups.

This pattern of results is clarified by looking at the distributions of the judgment values for the four conditions. As can be seen in the categorized histogram displayed in Fig. 2, the distribution of the number of observations for each possible judgment value strongly depends on the programmed condition. The distributions were slightly asymmetrical, and almost identical, for the two mixed-contingency conditions. In both cases, the most frequent value, was 0, the median value. In the high-contingency diagnostic task, the distribution is also normal-like, with most judgments around intermediate or high levels. This means that all subjects in that condition inferred some degree of contingency between the cues. Conversely, in the high-contingency predictive task (where the two cues were defined as potential causes), the distribution is clearly bimodal. Most subjects did not infer the existence of correlation between the two cues, and, thus, their judgments are around 0. However, a significant portion of the sample (6 out of 16 participants) gave very high between-cues correlation judgments. Kolmogorov–Smirnoff–Lilliefors tests

Fig. 2. Categorized histogram (number of observations/judgment value) for the four conditions in Experiment 2.
corroborated this interpretation. The judgment distribution was non-normal in the high-contingency/predictive group ($d = .269$), but it was normal in the high-contingency/diagnostic group.

Non-parametric Wald–Wolfowitz tests (using $p = .0125$, according to the Bonferroni error correction method), which are sensitive both to the location and shape of the distributions, showed significant differences between the predictive and the diagnostic high-contingency groups, $Z = -3.19$, and between the high-contingency and mixed-contingency diagnostic groups, $Z = -3.95$. Note that this pattern of results is identical to the one shown by parametric tests.

**General discussion**

The main results obtained in the present experiments can be summarised as follows. First, people do not obtain correlation knowledge exclusively from observed correlations, but also, indirectly, from complex scenarios in which several causal relationships share a common element, as demonstrated by the inference of mediated correlation in experiments 1a and 2. Second, naïve reasoners are more likely to infer the existence of correlation between two separately trained effects of a common cause than between two separately trained causes of a common effect. Most participants inferred the existence of correlation between two cues that had previously been defined as effects of a common potential cause (experiment 1a, and experiment 2, diagnostic condition), but not between two potential causes of the same effect (experiment 1b, and experiment 2, predictive condition). Third, this type of mediated correlation learning occurs, but with a significantly lesser intensity, in situations in which the two cues are known to be causes of a common effect (Experiment 2, predictive condition). However, the intermediate mean judgments observed in that condition reflect the existence of two separate samples of subjects. Most people (62.5%) do not infer mediated correlations at all, whereas approximately one third of the sample (37.5%) perceived mediated correlation to the same degree as those in the diagnostic (common-cause) condition.

This pattern of results is in partial agreement with causal learning theories that maintain that causal knowledge can be represented in a complex manner, including the directional features of causal relationships. Causal structures, or causal models, seem to determine, in a top–down manner, how non-observed correlations are inferred. However, some subjects do not show the expected difference, which can be interpreted either as a consequence of the operation of basic associative mechanisms (only in one third of the whole sample), or as a result of using models alternative to the common-cause and the common-effect schemata considered here—for example, a causal chain model (it is important to acknowledge that if such a model were applied, a high correlation between both causes would be inferred).

Associative or data-driven mechanisms cannot account for the effect of causal directionality obtained in this work. First and foremost, by no means can data-driven models explain the effect of causal directionality on mean between-cues correlation judgments. According to an associative account, the mechanism responsible for inferring mediated correlation would operate during the second phase of the
experiment, by reinforcing the connection between the mental representation of the cue presented during that second phase (a cue that is actually present) and the mental representation of the cue presented during the first phase (which is not present, but can be elicited both by the common outcome and the context). Thus, associative models predict the same amount of mediated learning in predictive (common-effect) and diagnostic (common-cause) situations. The only way to account for a difference between these two conditions would be to select different learning rate parameter values for each condition, which should be based on firm theoretical grounds. Even if we accept the possibility of different parameters, associative models would still need to face two difficulties. First, although different learning rates during stages one and two of the task would yield different levels of inferred correlations, at the same time, they would yield different estimates for the strength of the causal relations between each cue and the outcome, a prediction proved to be wrong. Second, associative models need to make post hoc extra assumptions to account for the bimodal distribution found in the high-contingency predictive condition of Experiment 2.

Actually, some models predict (i.e. Aitken, Larkin, & Dickinson, 2001; Van-Hamme & Wasserman, 1994) not only that the presence of a cue activates its mental representation but also that the mental representation of an absent, though expected, cue could be activated with a negative value (or directly in the A2 state, in terms of the reformulated SOP model proposed by Aitken et al.). The joint activation of the representations of an absent and a present cue is supposed to strengthen an inhibitory link between them. However, a positive inferred correlation reflects the existence of an excitatory link. Thus, in order to explain our data on inferred correlation, these models should be modified to include the possibility that absent-expected cues could be activated positively (this could be done by discriminating between explicitly absent cues and cues about which there is no information available).

Alternatively, it can be argued that most people construct causal models during learning, according to previous knowledge, and use them to infer or discard the existence of mediated correlations. The main feature of such models is that they cannot be boiled down to mere contingencies or mental associations, as they incorporate causal directionality. However, one out of three subjects did not behave in accord with the proposed models. These participants either could be learning in an associative manner or could be using models other than the common-effect and the common-cause models considered here. For example, in the high-contingency predictive condition of experiment 2, in which contingencies between each cue and the outcome are high and equal to each other, some participants could presume that one cue causes the other, which, in turn, would cause the outcome with a given probability. In this case, both cues would be considered to be maximally correlated.

On the other hand, there are several ways in which data-driven (i.e., associative) and cognitively-driven (i.e., causal-model based) theories can coexist. According to the simplest one, data-driven mechanisms and cognitively-driven mechanisms represent different learning strategies rather than alternative and mutually exclusive mechanisms. The fact that humans have both computational and associative capacities has been stressed in other areas of cognition. For example, in the field of language acquisition, Pinker & Prince (1988) formulated a two-part model, in which regular past tense forms
depend on a rule system, and irregular forms are associatively learned and produced. In a similar fashion, Markus, Vijayan, Rao, & Vishton (1999) have also recently proposed that human infants possess both associative-learning and rule-learning capacities. De Houwer & Beckers (2002) have recently made a similar proposal to account for the puzzling corpus of data available in the contingency learning literature.

From this point of view, in some situations subjects will use causal mental models to codify the information provided during the experiment, whereas in others they will adopt a bottom-up strategy and will codify and interpret the information in a somewhat “raw” form. Both task and individual features (permanent or situational) could determine the type of mechanisms to be applied in a given scenario. In our experiments, if this were the case, the higher the number of subjects applying causal models is, the more sensitive the influence of causal directionality on inferred correlations would be. But, at the same time, if a significant number of subjects do not build or apply causal models relevant to the presented scenarios, the mean difference between the predictive and the diagnostic conditions will be quantitative rather than qualitative. As observed in Experiment 2, inferred correlation would appear in both situations, but much more neatly in the diagnostic task. The bimodal distribution of between-cues correlation judgments observed in this experiment seems to support this interpretation.

To date, there is no information to determine whether people who infer mediated correlation in the predictive condition are behaving associatively or are using alternative models. However, our data point to the need for addressing such a question. Two factors have been mentioned as potential causes of an effect at this level. The first is the availability of explicit causal information at the beginning of the task that can guide the integration of contingency information in a causal structure or mental model (Waldmann & Martignon, 1998); the second is the availability of enough attentional and executive resources. Preliminary data indicate that complicating the task by using multiple cues and outcomes reduces the effect of causal directionality (Perales & Catena, 2001).

In summary, associative (or, in general terms, data driven learning mechanisms) cannot account for the effect of causal directionality on inferred correlation. Potentially, data driven models could account for the judgments of some subjects that do not seem to discriminate between predictive and diagnostic situations. However, whether this group of subjects behaves according to mental models, such as the causal chain model described previously, or they are using an associative rule to infer the existence of correlation between the cues is still an open question. Consequently, the present results support the idea that reasoners incorporate learning experiences into mental schemas that cannot be reduced to mere contingencies. Further research should isolate factors -such as executive load- that could determine which learning strategy is selected when the observed behavior deviates from this general trend.

Appendix A

The normative analysis of correlation inference after a mediated contingency learning procedure presented here, partially based on Cheng (1997), proposes that
causes have the inherent capacity to produce their effects and, under some assumptions, their existence can be inferred from correlation information.

In the case of the common-effect scenario (see Fig. 3A), three causal factors should be considered: the two candidate causes, \( C_1 \) and \( C_2 \), that can produce the common effect by themselves, and an alternative causal factor (\( A \)), or contextual compound, unknown for the observer. In Fig. 3, \( x \), \( y \), and \( z \), denote the causal powers of \( C_1 \), \( C_2 \), and \( A \), respectively. The causal power of a given potential cause \( C \) will be the probability of this causal factor producing the effect by itself, defined as the probability with which this factor would produce the occurrence of the effect in an ideal context in which all the other alternative causes of the same effect are absent. A necessary assumption is that a naive reasoner normally builds the simplest possible model compatible with the available information. By default, the noisy-or schema (see Pearl, 2000) will be applied. Accordingly, two potential causes of the same effect are considered independent of each other unless complementary information is provided. In our common-effect case, no normative conclusions about the possible correlation between the two known causes can be addressed from this model.

The situation seems to be rather different in the common-cause scenario (Fig. 3B). In this case, only two causal factors should be considered: the common cause \( C \), producing the effects \( E_1 \) and \( E_2 \) by itself with causal power \( x \) and \( y \), respectively; and the possible alternative cause \( A \), producing the same two effects by itself with causal power \( r \) and \( s \). According to this schema, a set of four probabilities would be necessary to deduce the level of between effects correlation, \( P(E_1 \land E_2) \), \( P(\neg E_1 \land E_2) \), \( P(\neg E_1 \land \neg E_2) \), and \( P(E_1 \land \neg E_2) \). However, according to Probability Theory, these probabilities can be deduced from a simpler set of probabilities, \( P(E_1) \), \( P(E_2) \), and \( P(E_1 \land E_2) \), in such a way that the degree of correlation between \( E_1 \) and \( E_2 \) can be derived from these probabilities without directly observing their correlation.

The probability of \( E_1 \) results from the combination of the causal power of the candidate cause and the possible alternative cause (\( x \) and \( s \) stand for their respective

---

Fig. 3. Schematization of the normative analysis provided for the two causal models hypothesised to underlie predictive (common-effect) and diagnostic (common-cause) tasks. Letters \( C \) and \( A \) stand for known and unknown (contextual) causes of the observed effects (represented by letter \( E \)). \( x \), \( y \), \( z \), \( r \), and \( s \) represent causal powers.
causal power, see Fig. 3A). According to the causal schema described previously where both cause are independent, this probability will be:

\[ P(E1) = P(C) \cdot x + P(a) \cdot r - P(C) \cdot P(a) \cdot x \cdot r, \]  
(A.1)

where \( P(E1) \) means the probability of the effect \( E1 \); \( P(C) \) and \( x \), stands for the probability of the candidate cause and its causal power and \( P(a) \) and \( y \), for the probability of the alternative cause and its causal power in the general population.

Similarly,

\[ P(E2) = P(C) \cdot y + P(a) \cdot s - P(C) \cdot P(a) \cdot y \cdot s, \]  
(A.2)

where \( P(E2) \) represents the probability of the second effect \( E2 \), and \( P(C) \) and \( y \) and \( P(a) \) and \( s \) represent the probability and causal powers of the candidate alternative causes to generate the effect \( E2 \).

The joint probability of both effects, \( E1 \) and \( E2 \) can be deduced as (see Fig. 3B),

\[
P(E1 \land E2) = P(C)xy + P(a)rs + P(a)P(C)xs + P(a)P(C)yr - P(a)P(C)ys - P(a)P(C)xr - P(a)P(C)xys + P(a)P(C)xrs + P(a)P(C)xyrs,
\]  
(A.3)

where all terms have been previously detailed.

Solving this system of equations requires six parameters. \( P(a) \), \( P(C) \), \( x \), \( y \), \( r \), and \( s \).

1. \( a \) is the background and it is supposed to be always present. Therefore, \( P(a) = 1 \).
2. \( x \) and \( y \) are the causal powers of \( C \) to produce \( E1 \) and \( E2 \), respectively. It is assumed that their value, as well as \( P(C) \), is computed during the learning task according to the experienced contingency.
3. \( r \) and \( s \) can also be estimated by means of computing the probability of the effect in those trials in which the cause \( C \) is absent, being \( r = P(E1 / \sim C) \), and \( s = P(E2 / \sim C) \).

Accordingly, the degree of correlation between \( E1 \) and \( E2 \) can be computed by giving values to every parameter in the model and more importantly, without the need of directly observing such correlation.

References


