

# High-Level Cognitive Processes in Causal Judgments: An Integrated Model

Andrea Stocco (stocco@units.it)

Daniilo Fum (fum@units.it)

Stefano Drioli (drioli@psico.units.it)

Dipartimento di Psicologia, Università di Trieste  
via S. Anastasio 12, I-34134, Trieste, Italy

## Abstract

The problem of whether human causal judgments could be better explained by associationistic or probabilistic accounts is dealt with in the paper that reviews the basic tenets of the power PC theory (Cheng, 1997), the most famous of the probabilistic explanations, and discusses some results obtained by Fum & Stocco (2003) that are at odds with power PC predictions. An integrated model is described that is capable of explaining those findings, and a new experiment is presented in which the predictions of the model and of the power PC theory are contrasted in the case in which the causal power of a compound cue is equal to one of its components. The results clearly corroborate the model that provides, moreover, an explanation for some data that lie outside the scope of the power PC theory.

## Introduction

Recent research on adult causal cognition has been focusing on two main kinds of theoretical explanations that capture many of the central findings in the field.

Associative accounts (Shanks, 1995) consider the causal reasoning performed by humans as similar to the classical conditioning happening in animals and claim that, because both processes involve the detection of the same predictive relations, they may use a common mechanism. The most famous model in this class is that of Rescorla & Wagner (1972)—henceforth R&W—that has been successfully applied to account for a series of phenomena—like blocking (Kamin, 1969), overshadowing (Price & Yates, 1993), conditioned inhibition (Chapman & Robbins, 1990), and contingency effects (Dickinson, Shanks, & Evenden, 1984), to name only a few—that were originally discovered in animals, and that have been demonstrated to play a critical role in human causal learning, too.

Probabilistic theories, on the other hand, rely essentially on the analysis of the contingencies that organisms are supposed to acquire by interacting with their environment, and try to estimate the extent to which a cue (or potential cause) can determine a given outcome. The most famous among these accounts is constituted by Cheng's power PC theory (Cheng, 1997), an extension of the probabilistic contrast model developed by Cheng & Novick (1990).

Fum & Stocco (2003) argued that associative and probabilistic models possibly cover distinct steps in human causal induction, with associative accounts describing the processes by which people (and animals) notice and extract statistical connections between events, and probabilistic models capturing the reasoning skills brought to bear in causal cognition.

Investigating the role of compound cues in causal judgments, however, they obtained experimental findings that could not be explained, in their entirety, by either group of theories.

In the paper we review the power PC theory and illustrate some results obtained in Fum & Stocco (2003) that seem to falsify it. We present a new model that, while being compatible with previous data, is able to explain those puzzling results. We describe an experiment in which our model and the power PC make contrasting predictions, and we present findings that corroborate our hypothesis. We discuss some further data that, while implied by our model, are out of the scope of the power PC theory. We conclude the paper by summarizing the features of our account of human causal cognition and by outlining some possible developments.

## A Probabilistic Account

Perhaps the simplest of the probabilistic models of causation is given by the  $\Delta P$  rule (Jenkins & Ward, 1965) that formalizes the idea that people mentally compare the frequency of an outcome  $O$  in presence and in absence of a given cue  $C$ :  $\Delta P_c = P(O|C) - P(O|\neg C)$ . If the difference is around 0, the outcome is just as likely when the cue is present as when it is absent; if it approaches 1,  $C$  is perceived as producing  $O$ ; if it approaches  $-1$ , the cue is seen as preventing the outcome.

Relying on this idea, Cheng & Novick (1990) developed their probabilistic contrast model assuming that, in presence of a set of possible causes for an effect, the  $\Delta P$  for each cause is computed on the so-called *focal set*, defined as “a contextually determined set of events that the reasoner uses as input to the covariation process” (Cheng, 1997, p. 371). When a putative cause is taken into account, all other causal factors are kept constant within the focal set, and  $\Delta P$  is computed on a background of constant alternative causes.

The transition from the probabilistic contrast model to the power PC theory was motivated by a series of problems that could not be adequately explained by the former nor by alternative associative accounts like the R&W. The power PC theory essentially computes how much a  $\Delta P$  judgment should be discounted for providing an estimate of the causal power of a cue. It also detects special conditions in which the causal power cannot be deduced from  $\Delta P$ .

One of the tenets of the theory is that, whenever the possible alternatives to a candidate cause  $C$  are kept under control and  $\Delta P$  is non negative,  $\mathbf{C}$  (i.e., the causal power of  $C$  to generate the outcome  $O$ ) is given by:

$$\mathbf{C} = \frac{\Delta P_c}{1 - P(O|\neg C)}$$

According to the power PC theory, identical values of  $\Delta P$  associated with different values of  $P(O|C)$ , the *base rate*, will lead to different causal judgments. When the alternative causes are controlled, the theory predicts that, with  $\Delta P$  kept constant, the causal power increases with an increase in the value of the base rate. As a special case, if the base rate is equal to 1, the causal power remains undefined, because the denominator becomes 0. If the base rate is equal to 0, the power PC reduces to the probabilistic contrast model, and the causal power depends exclusively on  $\Delta P$ . Finally, if  $\Delta P$  is 0, the causal power of  $C$  is 0, too.

Fum & Stocco (2003) focused on some interesting consequences of Cheng's theory concerning the role of compound cues, and set up an experiment to test them. To reduce the complexity of the theoretical framework, and to establish a clear experimental paradigm, four assumptions were made. First, the causal power of a generic cue  $A$  was defined as the probability that, all other things being equal, the cue would produce the outcome  $O$ :  $A = P(O|A)$ . Second, a given outcome had a null probability of being obtained in absence of the cue:  $P(O|\neg A) = 0$ . Third, all cues were considered as independent. Fourth, all the cues were pure causes: none of them was an enabling condition (Cheng & Novick, 1991) nor needed any enabling conditions to produce its effect.

Given these assumptions, it is possible to deduce<sup>1</sup> some important consequences from the power PC theory. We focus here on two of them:

**Irrelevance of Compound** Previous experience with a cue presented in a compound form should be irrelevant to the judgment of its causal power, given that there are trials in which the cue appears alone. It is a tenet of both the power PC theory and of the probabilistic contrast model that only items in the focal set—where everything, but the candidate cause whose causal power is being evaluated, is kept constant—are taken into account to compute  $\Delta P$ . Let us consider, for instance, the classical backward blocking paradigm (Chapman, 1991; Dickinson & Burke, 1996; Shanks, 1985), where compound trials of the form  $(A, B \rightarrow O)$  are followed by a set of trials of the form  $(A \rightarrow B)$ . In this context, an adequate focal set to evaluate the causal power of the blocking cue  $A$  is constituted by trials  $(A \rightarrow O)$  only, because by including  $(A, B \rightarrow O)$  in the set, the cue  $B$  would also vary. The power PC theory therefore predicts that a previous presentation of a compound cue  $(A, B \rightarrow O)$  should not influence the following judgment for cue  $A$ .

**Equalization to Compound** Sometimes it could be necessary to estimate the causal power of a cue over an inadequate focal set. Taking the example of backward blocking again into account, it should be noted that the trials  $(A, B \rightarrow O)$  constitute an inadequate focal set for evaluating the causal powers of  $A$  and of  $B$  because both cues are covariant within the same set. However, this is exactly what participants in the control group of that paradigm are requested to do, and what a theory is supposed to provide an explanation for. When participants are forced to make a judgment, they should adopt the trials  $(A, B \rightarrow O)$  as a focal set, and this would lead them to assign both cues the same causal power of the compound.

Given the fact that the effect is never obtained without the cause, and that each possible cause appears in the set of trials  $(A, B \rightarrow O)$ , it is possible to demonstrate that  $A = P(O|A, B)$  i.e. the causal power of cue  $A$  should be equal to the probability of obtaining the outcome given the compound. The same should be true for  $B$ .

The experiment carried out in Fum & Stocco (2003) obtained findings that falsified these predictions. More precisely, contrary to the irrelevance of compound hypothesis, judgments concerning a cue  $A$ , experienced only in a compound form (i.e. together with another cue  $B$ ), were significantly *higher* than judgments for the same cue experienced alone. In a similar vein, and contrary to the equalization to compound prediction, the judgments for a cue  $A$ , experienced only in a compound form, were significantly *lower* than judgments given to the compound cue embedding  $A$ .

While no theoretical explanation for these results was provided in the paper, the findings clearly suggested the existence of important factors determining causal judgements that lie beyond the scope of the power PC theory.

## An Integrated Model

Trying to find an explanation for the results reported in Fum & Stocco (2003), we assume that people are able to acquire some knowledge about the contingencies that exist between cues and outcomes. A significant role is played in this phase by associative processes that contribute to the construction of an internal representation for the magnitude of the (single and compound) cues that were directly experienced. There is evidence that it is possible to spontaneously learn such knowledge by interacting with the environment (e.g., Hasher & Zacks, 1984), and we assume that people rely on this information in providing the judgments for those situations they actually encountered.

When a judgment about the causal power of a cue experienced only in compound form is required, the information about the stored magnitudes is used to infer the causal power of the individual novel stimuli, too. This process resembles reverse engineering, because people are supposed to figure out a conceivable distribution for the magnitudes of the single cues that could originate the magnitude of the compound representation.

Let us consider the top panel of Figure 1, that depicts the situation typically encountered in a blocking paradigm. The model assumes that, by interacting with the environment and by noticing the contingencies between cues and outcomes, people are able to construct an internal representation for the causal power of the cues they experience directly—for instance,  $A$ ,  $(A, B)$ , and others, like  $C$ . The stored magnitude could be a more or less faithful representation of the actual causal power of a given cue but, in any case, it constitutes the basis for causal judgments. To estimate the causal power of an experienced cue, people rely on its magnitude representation, and translate it into the required numerical scale.

When requested to provide an estimate for a cue that was experienced only in compound form ( $B$ , in our example), they try to figure out a sensible value for it—in our case, a magnitude for  $B$  compatible with the magnitude of both  $A$  and the compound  $(A, B)$ . This process involves a comparison only between those cues that are relevant for deriving the causal

<sup>1</sup>We refer to the original paper for the mathematical derivations.

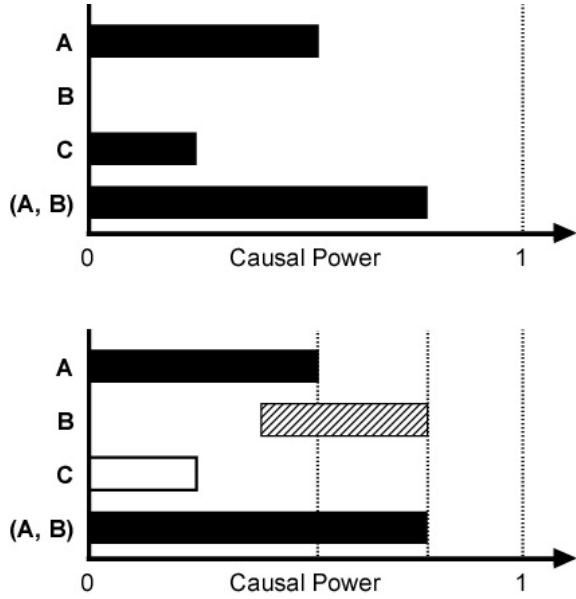


Figure 1: Phases of causal judgement. *Top (associative)*: Causal powers for cues represented as inner magnitudes. *Bottom (probabilistic)*: Inferring the value of cue  $B$ , that has not been experienced.

power of  $B$  (in our example  $A$  and  $(A, B)$ ) and excludes the others (in our case,  $C$ ). The set of cues taken into consideration conforms to the notion of focal set.

First, we shall observe that the causal power of  $B$ , which we denote through its boldface name  $\mathbf{B}$ , cannot be smaller than the difference between  $(\mathbf{A}, \mathbf{B})$ , the causal power of the compound, and  $\mathbf{A}$ : if it were so, a certain part of the overall compound effect would remain unexplained: a “rod” shorter than  $(\mathbf{A}, \mathbf{B}) - \mathbf{A}$  could not cover the whole length of the rod representing  $(\mathbf{A}, \mathbf{B})$ . Therefore,  $\mathbf{B}_{min} = (\mathbf{A}, \mathbf{B}) - \mathbf{A}$ . Generally, some part of the causal power of  $B$  will be shadowed by  $A$ : if we simply subtract  $\mathbf{A}$  from the compound  $(\mathbf{A}, \mathbf{B})$ , we would in fact grossly underestimate the causal power of  $B$ . On the other hand,  $\mathbf{B}$  cannot be greater than  $(\mathbf{A}, \mathbf{B})$ , so that,  $\mathbf{B}_{max} = (\mathbf{A}, \mathbf{B})$ .

The model therefore assumes that is “rational” to provide as a judgment for the causal power of  $B$  a value lying between  $\mathbf{B}_{min}$  and  $\mathbf{B}_{max}$ . All the values between this range are plausible and coherent with the magnitude of the associatively experienced contingencies. The particular judgments provided by participants vary stochastically between this range. The mean expected value for  $\mathbf{B}$  is therefore obtained by weighting each possible  $\mathbf{B}$  by its probability  $P(\mathbf{B})$ :

$$\bar{\mathbf{B}} = \int_{\mathbf{B}_{min}}^{\mathbf{B}_{max}} \mathbf{B} P(\mathbf{B}) d\mathbf{B}$$

For any symmetrically distributed probability function  $P(\mathbf{B})$ , the previous equation reduces to the average between  $\mathbf{B}_{min}$  and  $\mathbf{B}_{max}$ :

$$\bar{\mathbf{B}} = (\mathbf{A}, \mathbf{B}) - \frac{1}{2} \mathbf{A}$$

## Explaining Previous Results

Not surprisingly, the model can accommodate the results obtained by Fum & Stocco (2003). Two main findings were reported in that paper. First, some associative effects resulted in a systematic distortion of the causal judgments provided by the participants. The model assumes that these effects are confined to the first phase of the process leading to causal judgments, where inner magnitudes of contingencies are supposed to be acquired.

The second result is more interesting, and it seems critical for the power PC theory. In order to account for backward blocking—one of the most robust and popular contingency learning phenomena—a theory should be able to explain how people make a causal judgment about a cue that has been experienced only in compound form. As previously noted, power PC either should exclude taking into account the inadequate focal set  $(A, B \rightarrow O)$ , denying thus itself the possibility to account for backward blocking, or should predict, by using only that available set, that the judgments about the causal power of  $A$  and  $B$  will be equal that of the compound  $(A, B)$ .

Our model makes a different prediction. According to it, participants are supposed to construct a mental representation of the causal power of  $A$  and  $B$  such that, by joining (and possibly overlapping) them, they will cover that of the compound  $(A, B)$ . Because the magnitude of the causal power of one of the cues, let us say  $\mathbf{A}$ , should be obviously comprised between 0 and  $(\mathbf{A}, \mathbf{B})$ , the estimate for the mean causal power  $\bar{\mathbf{B}}$  could be computed by averaging on the predicted values of  $\mathbf{B}$ , computed on all the values for  $\mathbf{A}$  comprised between these extremes:

$$\bar{\mathbf{B}} = \int_0^{(\mathbf{A}, \mathbf{B})} \left( (\mathbf{A}, \mathbf{B}) - \frac{1}{2} \mathbf{A} \right) d\mathbf{A} / (\mathbf{A}, \mathbf{B})$$

By solving this equation we obtain:

$$\bar{\mathbf{B}} = \frac{3}{4} (\mathbf{A}, \mathbf{B})$$

The same result will hold, of course, for  $\bar{\mathbf{A}}$ .

In the experiment of Fum & Stocco (2003) the value for  $(\mathbf{A}, \mathbf{B})$  was set to 0.80. Under this condition, the model predicts that  $\mathbf{A} = \mathbf{B} = 0.60$ . The judgments provided by participants were  $\mathbf{A} = 0.62$  and  $\mathbf{B} = 0.61$ , respectively, with the difference being not statistically significant. It is useful to remind that, according to the power PC theory  $\mathbf{A} = \mathbf{B} = (\mathbf{A}, \mathbf{B}) = 0.80$ .

## Some New Predictions

A model should be considered as good not because it is able to explain previous data but because it allows making bold predictions about future events. For most of the cases, our model produces estimates of the causal power that are close to those provided by the power PC theory. It makes, however, a completely different prediction when, in an extreme blocking situation, the causal power of the compound is equal to the causal power of one of its components:  $(\mathbf{A}, \mathbf{B}) = \mathbf{A}$ .

**Causal Judgments** Under this condition, the power PC theory predicts that, independently of the values assumed by  $(A, B)$  and  $A$ , the cue  $B$  will be perceived as having a null

causal power. In this case,  $B$  is the candidate cause, and  $A$  is a background cause, as assumed in Cheng's framework. The focal set for  $B$  is constituted by all of the trials ( $A, B \rightarrow O$ ) and ( $A \rightarrow O$ ). In this set,  $P(O|B)$  may be estimated as  $P(O|A, B)$ , and  $P(O|\neg B)$  is given by  $P(O|A)$ . Given that, we can apply the standard equation for the causal power:

$$\mathbf{B} = \frac{P(O|A, B) - P(O|A)}{1 - P(O|A)}$$

Since  $P(O|A, B) = P(O|A)$ , it follows that  $\mathbf{B} = 0$ .<sup>2</sup>

On the contrary, according to our model,  $\mathbf{B}$  will be given a value equal to  $(\mathbf{A}, \mathbf{B}) - \mathbf{A}/2$ . Because,  $(\mathbf{A}, \mathbf{B})$  has been supposed equal to  $\mathbf{A}$ , it follows that  $\mathbf{B} = \mathbf{A}/2$ , i.e., we expect that the causal power of  $B$  will be judged to be about half of the value of the causal power of  $A$ .

Because the model provides some details about the processes underlying causal judgment, it allows making some predictions that lie outside the scope of competitive accounts. More particularly, it predicts the following:

**Confidence Ratings** According to the model, trying to provide a judgment about the causal power of an experienced cue (e.g.  $A$  in the backward blocking paradigm), participants rely on an existing stored representation. On the other hand, when requested to assess the causal power of a cue that has been experienced only in a compound form (e.g.,  $B$ , in the same paradigm) they cannot access a similar representation to assign a reliable value to  $\mathbf{B}$ , and that constitutes an important source of uncertainty. When requested to estimate the confidence according to which they provide their causal judgments, participants should therefore trust their judgment for cue  $A$  more than that provided for  $B$ .

**Latencies** In the backward blocking paradigm—in which the presentation of a compound stimulus ( $A, B \rightarrow O$ ) is followed by the presentation of one of its components, e.g., ( $A \rightarrow O$ )—the judgment concerning the causal power of  $A$  could be produced by reading off the value of its internal representation built according to associational principles. To provide a judgment for  $B$ , on the other hand, it is necessary to take into account the range of possible values that a coherent judgment could assume, i.e., it is necessary to resort to the second phase hypothesized by the model. As a consequence, the time need to provide a judgment for  $\mathbf{B}$  should be longer than that spent in trying to assess the value for  $\mathbf{A}$ .

## The Experiment

To put these ideas under empirical testing, we carried out an experiment, using the Tanks paradigm introduced by Shanks (1985), in which the predictions of our model were directly compared with those deriving from the power PC theory.

### Method

**Participants** The participants were 111 college students (28 males and 83 females) aged between 18 and 36 years

<sup>2</sup>Cheng (1997) derived a computational analysis of the R&W model showing that, under particular conditions—that were met by our experiment—it asymptotically computes a probabilistic contrast over a focal set. It is therefore possible to conclude that, the R&W model makes here the same prediction of the power PC, i.e.,  $\mathbf{B} = 0$ .

(mean and median = 20) enrolled in an introductory Psychology course.

**Design and Procedure** Participants saw a series of trials in which a picture of an army tank moved across a computer screen. On every trial a weapon system fired, and the tank was hit by one or two projectiles; in some trials the tank was destroyed, in others it remained undamaged. At the moment the weapon fired, one or two colored lights went on in the lower part of the computer screen. The color of the light indicated the kind of projectile that was used. Conceptually, each light could be considered as a separate cue, and the explosion of the tank could be regarded as the outcome.

The experimental session consisted of two sets of 20 trials each. In every trial from the first set two projectiles were contemporaneously fired—this phase could be indicated by ( $A, B \rightarrow O$ ). In each trial from the second set one of the projectiles was fired alone ( $A \rightarrow O$ ). Three experimental conditions were set up, each condition differing in the probability of the tank being destroyed by the projectiles. The probabilities were equal to 0.2 (Low), 0.5 (Medium) and 0.8 (High). In the Low condition, therefore, the tank was (randomly) destroyed 4 times in 20 trials, while in the Medium and High condition it was destroyed 10 and 16 times, respectively. The probability of the tank being destroyed by two projectiles in trials from the first set was the same as the probability of being destroyed by a single projectile in the others, i.e.,  $P(A, B \rightarrow O) = P(A \rightarrow O)$ . Finally, trials from the two sets were randomly interleaved for each participant, and participants were randomly assigned to an experimental condition.

Participants were requested to judge the efficacy of each kind of projectile on a scale ranging from 0 to 100, where 0 indicated null efficacy (i.e., the projectile never destroyed the tank) and 100 maximum efficacy (i.e., the projectile always destroyed the tank). They were also asked to indicate, on a seven point scale, the confidence with which they formulated their judgments about the causal power of each projectile. The last main dependent variable that was recorded was constituted by time needed to provide each causal judgment.

At the beginning, participants read an instructions sheet, written in Italian, that explained the task. After that, they saw four practice trials. The tank was randomly destroyed in two of the trials, and in the remaining two it was left undamaged.

At this point the experiment could start. Two colors (chosen in a set that comprised red, yellow, green, and blue) were randomly assigned to the two projectiles used in each experiment session, and the participants were exposed to the 20 trials of the first phase and 20 trials of the second one. To ensure that participants paid attention to the presentation trials, during the experiment four “control” screens appeared at randomly chosen times asking participants to indicate what they had just seen, i.e. which projectile/s was/were fired and whether the tank had been destroyed.

At the end of the presentation trials, participants were asked to provide their judgment about the efficacy of each of the two projectiles they had experienced. After that, they were requested to rate their causal judgments, i.e., to indicate how confident they were about the correctness of their answers.

**Stimuli and Apparatus** The experiment was performed on a PC equipped with a 15” LCD flat screen and headphones. A

custom-made program written in Java was utilized to present the stimuli and to record the participants' judgments. During the presentation trials, the picture of a tank (120 x 45 pixels) moved at constant speed crossing the screen from right to left. A disk (with a diameter of 300 pixel) in the center of the screen simulated the view finder of the weapon system and displayed a desert landscape. The area of the screen outside the disk was kept blank. The tank was visible only when it crossed the disk (employing 3300 ms to cover its diameter), in the remaining time participants could only hear the engine sound through the headphones.

When the tank was approximately at half of its path, completely visible within the view finder, the weapon fired: one or two gunfire sounds were heard and one or two lights, represented by round LEDs (with diameter of 150 pixels) were lit up with the color of the projectile that had been shot.

The tank was always hit, and 1000 ms after the LEDs were brightened, it flashed for 300 ms to simulate the projectile impact. In the trials in which the tank was destroyed, an explosion sound was heard, and the tank was covered by a dust cloud that, after it dissolved, left visible only the wreck. In the trials in which the tank was left undamaged it continued its course until disappearing from the view. In both cases the LEDs remained lit. Each trial lasted approximately 7.5 s; after that, with a shutter effect, the view finder was closed and opened again, and a new trial began.

The control screen utilized to monitor the participants' attention had four LEDs placed at the vertices of an imaginary rectangle positioned at the center of the monitor, each LED associated with two radio button labeled "Yes" and "No", respectively. Participants were asked to indicate which LEDs were lit (and which projectiles were fired) in the very last trial. Moreover, they had to indicate whether the tank had been destroyed or not by choosing between two more yes/no buttons.

The judgments about the efficacy of each projectile were collected through separate screens. In each screen a colored LED was presented together with a request to provide a judgment about the projectile by setting a slider. The mark was positioned at the middle of the slider and the value for the judgment was set to "unassigned". As soon as the participant started moving the mark, an integer value appeared on screen indicating the mark position on a scale ranging from 0 to 100. The confidence rating were collected by having participants check one of seven radio buttons. The buttons at the extremes were labeled with the Italian equivalents of "No Confidence" and "Complete Confidence", respectively.

## Results

To avoid considering data that did not accurately reflect the phenomena under investigation, participants that made four or more errors (over a total of 20 possible answers) in the control task were excluded from the sample.<sup>3</sup> The data of 19 (out of 111) participants were thus discarded, and the following analyses were carried out on the remaining ones: 28 participants in the Low condition, 30 in the Medium, and 34 in the High condition, respectively.

<sup>3</sup>The same criterion had been adopted in Fum & Stocco (2003).

Table 1: Mean causal judgments for **A** and **B**

	Low	Medium	High
Judgment for <b>A</b>	41.32	64.53	75.44
Judgment for <b>B</b>	23.48	39.83	48.53

**Causal Judgments** The causal judgments provided by participants are reported in Table 1 and illustrated in the top panel of Figure 2. A mixed-design ANOVA was carried out having Condition (Low vs Medium vs High) as a between-subjects and Judgment (A vs B) as a within-subjects variable.

The analysis showed as significant the main effects of the Condition ( $F(2, 89) = 34.44, MSE = 1125.90, p < .0001$ ) and of the Judgment ( $F(2, 89) = 55.29, MSE = 24357, p < .0001$ ) but not their interaction. Contrary to the power PC predictions, participants provided judgments for the causal power of *B* that were completely different from the expected zero value ( $t(88) = 14.50, p < .0001$ ). In accordance with the predictions of our model, their judgments for the causal power of *A* and *B* differed significantly, and the value of the judgments increased with an increase in causal power of the compound stimulus (*A, B*). The model, however, makes a stronger prediction, i.e., that the ratio between the two stimuli should be constant. We calculated this ratio for each participant, and then computed the mean of the ratios for each condition. Results are reported in the bottom part of Figure 2: ratios remained constant across conditions, with only slight and insignificant differences among them. Values of the ratios were around 0.6, close to our estimate, i.e. 0.5. This result could be considered more than satisfactory being our model completely parameter-free.

**Confidence Ratings** In analyzing confidence ratings and latencies, we pooled the data of all the participants because

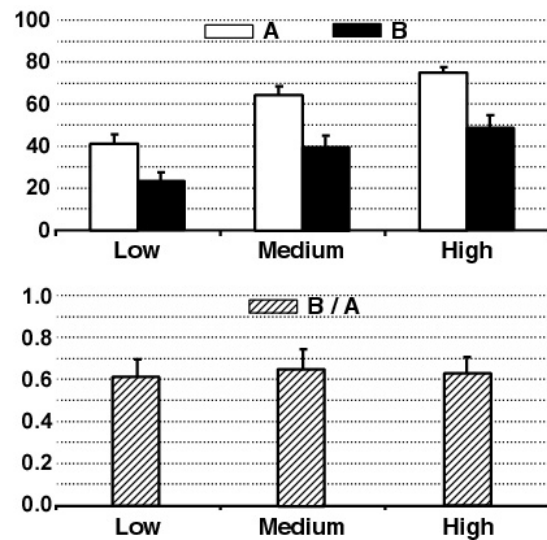


Figure 2: Mean causal judgments (*top*) and mean ratios between **B** and **A** (*bottom*) in the experimental conditions.

our model does not discriminate, under these aspects, among the different conditions.

We found that the confidence ratings for **A** (mean = 3.85) were indeed greater than those for **B** (mean = 3.52), the difference—as revealed by a two-tailed, paired *t*-test—being statistically significant ( $t(91) = 2.48, p = 0.01$ ). However, to take into account the possibility that the confidence ratings—collected through a seven point Likert scale—did not conform to a normal distribution, we conducted also a Wilcoxon Matched-Paired test that confirmed the existence of the effect ( $T = 676.00, N = 92, p = 0.04$ ).

**Latencies** An even strongest corroboration for our model came from the analysis of latencies. As predicted, the mean time needed to express a judgment for *A* (17.18 s) resulted smaller than the time needed for *B* (21.31 s). A *t*-test confirmed that the difference was significant ( $t(90) = 2.14, p = 0.04$ ). The difference remained significant taking into account the square root ( $t(90) = -2.11, p = 0.03$ ) and the logarithm ( $t(90) = 2.07, p = 0.04$ ) of the latencies. Causal induction processes seem to respect the time course we hypothesized.

## Conclusions

In the paper we presented a model of high level cognitive processes in causal induction that is able to explain previous findings that resulted antithetical to some predictions of the power PC theory, and that can take into account new data that are at odds with, or beyond the scope of, that theory. The model assumes that, when required to provide a causal judgment, people recur to both associative and probabilistic processes. These processes play, however, a different role in causal cognition: associative processes contribute to the construction of an internal representation of the power of directly experienced cues, while probabilistic reasoning is required to estimate the magnitude of the non directly perceived ones.

In the paper we have gone one step further in the description of the cognitive processes underlying such judgments, and we have extended the set of data that may be taken into consideration to discriminate between different accounts. We find particularly important the fact that participants were able to express faithful subjective confidence about their own ability to estimate the causal power of different cues, an indication that these estimates lie above the subjective threshold (Dienes & Perner, 1999) and, therefore, that some kind of explicit knowledge is required to provide these judgments.

We are currently working to extend the model on paradigms other than blocking, and to provide finer estimates of human causal judgment. In particular we think that associative effects, probably reflecting an evolutionarily older, non-specific learning system, shall be further investigated. Our model might be of guidance in determining under which conditions such effects may overcome the explicit processes we described.

## Acknowledgements

We thank Fabio Del Missier for some helpful suggestions and the fruitful discussions. We would also like to thank an anonymous reviewer for the inspiring comments.

## References

- Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 837–854.
- Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, *23*, 510–524.
- Cheng, P. W. (1997). From covariation to causation. A causal power theory. *Psychological Review*, *104*, 367–405.
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*, 545–567.
- Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, *40*, 83–120.
- Dickinson, A., & Burke, J. (1996). Within compound associations mediate the retrospective reevaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, *49B*, 60–80.
- Dienes, Z., & Perner, J. (1999) A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*, 735–755.
- Dickinson, A., Shanks, D. R., & Evenden, J. L. (1984). Judgment of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology*, *36A*, 29–50.
- Fum, D., & Stocco, A. (2003). The role of compound cues in causal judgment: Associative and probabilistic effects. In F. Schmalhofer, R. M. Young & G. Katz (Eds.), *Proceedings of EuroCogSci 03*, Mahwah, NJ: Lawrence Erlbaum, 127–132.
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *12*, 1372–1388.
- Jenkins, H., & Ward, W. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, *7*, 1–17.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton.
- Price, P. C., & Yates, J. F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition*, *21*, 561–572.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp.64–99). New York: Appleton-Century-Crofts.
- Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology*, *37B*, 1–21.
- Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge, UK: Cambridge University Press.