

The Role of Compound Cues in Causal Judgment: Associative and Probabilistic Effects

Danilo Fum (fum@units.it)

Department of Psychology, University of Trieste
Via S.Anastasio 12, I-34134, Trieste, Italy

Andrea Stocco (stocco@units.it)

Department of Psychology, University of Trieste
Via S.Anastasio 12, I-34134, Trieste, Italy

Abstract

Contrasting associative and contingency based explanations of adult causal judgment, we derived three new predictions from the power PC theory (Cheng, 1997) concerning the role of compound cues, and we tested them in a context in which an effect could be produced by independent cues considered only as possible causes and not as enabling conditions. The first prediction, independence from the compound, was confirmed, and corroborated the theory. The second and third prediction, concerning the irrelevance of, and the equalization to a compound cue, were falsified by the experimental results. These findings suggest the existence of important factors determining causal judgements that lie beyond the scope of power PC theory.

Introduction

The capability of learning causal relations constitutes a crucial feature of intelligent behavior. Both animals and humans extract causal information from the environment, and they are able to detect predictive associations among events. How do they process such information? This question has received two answers in cognitive science.

The first claims that causal learning is fundamentally an associative phenomenon that underlies both contingency judgments provided by humans and classical conditioning found in animals (Shanks, 1995). The second conceives it as a form of probabilistic reasoning. Several statistical models, that rely essentially on the analysis of contingency tables (Cheng & Novick, 1990; Anderson & Sheu, 1995), have been developed within this framework trying to assess the extent to which a cue (or potential cause) C can determine an outcome O . Perhaps the simplest, and most famous, of such models is the so called ΔP rule (Jenkins & Ward, 1965) that formalizes the idea that people mentally compare the frequency of the outcome in the presence and in the absence of the cue: $\Delta P_c = P(O | C) - P(O | \neg C)$. 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.

Within the associative tradition, the model developed by Rescorla & Wagner (1972) (henceforth R&W) asymptotically computes a variant of ΔP , provided that some limiting conditions hold. Originally developed to explain how the associative strength between an unconditioned and a conditioned stimulus changes with time, the model has been applied to causal induction by equating the conditioned stimulus to a cue C , and the unconditioned stimulus to an outcome O .

The R&W model states that, every time C or O are experienced, the associative strength $V_{C,O}$ existing between them is updated, as a consequence of the presentation trial, of a quantity

$$\Delta V_{c,o} = \mathbf{a}\mathbf{b}(\mathbf{I} - \sum_s V_{s,o})$$

where \mathbf{a} and \mathbf{b} are parameters depending on the salience of C and O , respectively, \mathbf{I} is the asymptotic value for $V_{C,O}$ while $\sum V_{s,o}$ constitutes the set of cues for O (including C) present in the trial. More concretely, \mathbf{I} can be interpreted as the trial outcome: if the trial results in O the value of \mathbf{I} is typically set to 1, it is set to 0 in the contrary case. In the same vein, $\sum V_{s,o}$ designates the expected outcome deriving from the set of possible causal factors. If the discrepancy between \mathbf{I} and $\sum V_{s,o}$ is huge, also $\Delta V_{C,O}$ will be huge; as far as the difference between them decreases, $\Delta V_{C,O}$ will decrease, too. Eventually, after a set of trials carried out under special conditions, like the so-called *nesting* (Cheng, 1997), the discrepancy between the expected and the obtained outcome will asymptotically reach 0. It is important to note that, differently from ΔP , in R&W the symbol $\Delta V_{C,O}$ does not indicate a measure of causal strength but a *variation* in the strength $V_{C,O}$ as a consequence of the presentation trial.

Researchers working with the R&W model typically distinguish between two versions of the model, according to the assumptions made about the \mathbf{b} parameter. If \mathbf{b} is assumed constant both in the trials in which O is present and in which it is absent, it is common to talk of a

restricted R&W model. If the parameter is allowed to assume different values, the R&W model is called *unrestricted* (Lober & Shanks, 2000). Chapman & Robbins (1990) demonstrated that the restricted model converges in its asymptotic predictions to ΔP in the special case in which C is unique.

In the 80s the R&W got the acme of its popularity. Attracted from the similarity between cue and outcome, on one hand, and cause and effect, on the other, researchers were trying to understand human causal cognition through associative principles. The R&W was thus applied to explain a series of phenomena found in animal behavior that were replicated in humans, too: from blocking (Shanks, 1985), to overshadowing (Spellman, 1996), from conditioned inhibition (Chapman & Robbins, 1990), to contingency effects (Wasserman, Elek, Chatlosh, & Baker, 1993).

A problem for the R&W and for contingency models based on the ΔP rule is given by the fact that covariation does not imply causation: we experience many things that are regularly associated without being led to infer the existence of a causal relation between them. For instance, every time I go out with an umbrella, the streets are wet; nobody would think, however, that the fact the pavement is slippery is caused by my decision to take a walk. This example illustrates one of the conditions that must hold to be able to infer a causal relationship, i.e., the capability to control for possible alternative causes (in this case, rain). To explain this and other more intriguing situations, Cheng (1997) proposed the so-called power PC theory (more precisely: the causal power theory of the probabilistic contrast model), considered nowadays as the more interesting and complete explanation of human causal judgment.

As it is implied by the name, the power PC theory constitutes an extension of the probabilistic contrast model by Cheng & Novick (1990). The model assumes that, in presence of a set of possible multiple causes for an effect, the ΔP for each putative 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 possible cause is taken into account, all other causal factors are kept constant within the focal set, and ΔP is computed on a background of constant alternative causes.

The transition to the power PC theory was motivated by a series of problems that could not be adequately explained neither by the R&W nor by the probabilistic contrast model. The power PC theory essentially computes how much a ΔP judgment should be discounted for providing an estimate of the causal power of a cue. It also detects special conditions when causal power cannot be deducted from ΔP .

One of the consequences of the theory is that, whenever the possible alternatives to a candidate cause C are kept under control and ΔP is non negative, $p(C)$ (i.e., the

causal power of C to generate the outcome O) is computed according to the equation:

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

(There exists a different formula to establish the causal power of C to *prevent* O , that is used when $\Delta P < 0$. However, we will not deal with it here.)

While contingency models and the (restricted) R&W assume that the only source for causality judgments is given by the evaluation of contingencies between the possible cause and the effect (i.e., the ΔP rule), the power PC theory predicts that identical values of ΔP associated with different values of $P(O|\neg C)$, the *base rate*, could result in different causal judgments.

When the alternative causes are controlled, the theory predicts that, with ΔP constant, the causal power increases with an increase in the value of the base rate. 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 ΔP . Finally, if ΔP is 0, the causal power of C is 0, too.

Several investigators tried to compare the power PC theory against the probabilistic contrast model and the R&W (e.g., Buehner & Cheng, 1997; Lober & Shanks, 2000). The power PC theory is able to explain some special conditions under which it is impossible to draw any conclusion about the causal power of cues, even if both the ΔP rule and the probabilistic contrast model would predict a positive causal effect. If the associative models constitute a bridge connecting conditioning and causal learning, the power PC theory connects the latter with scientific and experimental reasoning (Cheng, 1997).

In fact, none of the proposed approaches is capable of providing a complete and satisfactory account of causal cognition. Associative models try to explain how the links between causes and effects can be built and strengthened. According to the terminology of Marr (1982) they are algorithmic models that allow the making of precise predictions but that, on the other hand, are applicable only to sequential tasks developing within time, and it is sometimes difficult to figure out what they really compute, i.e. to consider them at the computational level. Cheng's proposals, (the probabilistic contrast model and the power PC theory) are defined only at the computational level, and they do not care about the underlying processes. It is important to highlight, moreover, that the predictions from Cheng's equations are only ordinal (Buehner & Cheng, 1997, p. 56).

While we appreciate the elegance of the power PC theory, we strongly feel that it lacks some of the descriptive power of associative models. We think that the theory needs to be integrated with work done in related areas such as human learning and memory, and to take advantage from research carried out on behavioral judgment and decision making.

Some predictions from the power PC theory

As a first step toward an integrated approach, we focussed on three, not obvious, consequences of Cheng's theory concerning the role of compound cues, and we set up an experiment to test them. To reduce the complexity of the theoretical framework, and to establish a clearer experimental paradigm, we made four assumptions.

First, the causal power of a generic cue A is defined as the probability that, all other things being equal, the cue will produce the outcome O : $p(A) = P(O | A)$. Second, a given outcome has a null probability of being obtained in absence of the cue: $P(O | \neg A) = 0$. In many cases, this will reduce the power PC theory to the probabilistic contrast model, and will rule out some disturbing effects found, for instance, by Lober & Shanks (2000). Third, all cues are considered as independent. This assures that, given cues A and B , the probability that the outcome will be obtained as a result of their joint presence can be computed quite simply through the formula

$$P(O | A, B) = P(O | A) + P(O | B) - P(O | A)P(O | B)^1$$

Fourth, all the cues are pure causes: none of them is an enabling condition (Cheng & Novick, 1991), or needs any enabling condition to produce its effect.

Given these assumptions, it is possible to derive three important consequences from the power PC theory.

Independence from compound: The judgment about the causal power of a cue is sometimes independent from the cue it is paired with. Let us consider, for instance, the standard backward blocking paradigm where, after a series of trials with two compound cues producing a given outcome ($A, B \rightarrow O$), only one of the cues is presented ($A \rightarrow O$). In this condition, the causal power of B is computed over a focal set comprising both series of trials, where A is kept constant while the candidate cause B varies. Because B is never presented alone, $P(O | B) = P(O | A, B)$ while $P(O | \neg B)$, reduces to $P(O | A)$. As a consequence, the causal power for B , is computed through

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

Because, according to the assumptions, cues A and B are independent, we can replace $P(O | A, B)$ with the formula discussed in our third assumption obtaining:

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

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

$$p(B) = P(O | B)$$

i.e., the causal power of cue B depends on B alone. This prediction could be tested by contrasting the judgment

¹ Such an equation is quite intuitive, and asserts that the probability of obtaining O , given both A and B , is the sum of the probability of O being generated by A or B minus the probability of being generated by both. A more formal demonstration is possible but not trivial, and we omitted it due to space limitations.

given to B , presented in a compound, with that given to another cue, with identical causal power, presented alone.

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 it 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 are taken into account to compute ΔP . In a focal set everything is kept constant, except the candidate cause whose causal power is being computed. Let us consider the backward blocking paradigm again. In this context, an adequate focal set to evaluate the causal power of cue A is constituted by trials $A \rightarrow O$ only, because in trials $A, B \rightarrow O$ the cue B is also varied. According to this consideration, it is possible to predict that a previous presentation of a compound cue $A, B \rightarrow O$ should not influence the following judgment given to cue A alone, in comparison to the judgment given to another cue, having identical causal power of A , but always presented alone.

Equalization to compound: Judgments about a cue experienced only in a compound stimulus are sometimes made equal to the causal power of compound. It is unclear what prediction the power PC theory makes in cases in which it is necessary to evaluate the causal power of a cue over an inadequate focal set. For instance, in the backward blocking paradigm, again, the trials $A, B \rightarrow O$ alone constitute an inadequate focal set for evaluating both the causal power of A and of B , since both cues vary within the set. However, this is exactly what participants in that paradigm's control group usually experience, and that a theory is supposed to provide an explanation for. When participants are forced to make a judgment, they should adopt $A, B \rightarrow O$ 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 $A, B \rightarrow O$, we obtain:

$$p(A) = \frac{P(O | A) - P(O | \neg A)}{1 - P(O | A)}$$

$$p(A) = \frac{P(O | A, B) - 0}{1 - 0}$$

$$p(A) = P(O | A, B)$$

i.e. the causal power of cue A is equal to the probability of obtaining the outcome given the compound. The same is true for B .

Since the causal power of a compound stimulus is larger than that of any of its component cues, we should obtain that the judgment given to cue A (or B) experienced as a compound stimulus, should be larger than the judgment given to another cue, having identical causal power of A (or B) presented alone.

To test all these predictions we carried out the following experiment.

Experiment

In the experiment participants saw three series of trials on which a picture of an army tank moved across a computer screen. The tank was seen through the view finder of a weapon system. In every trial the tank was hit by one or two projectiles of a different kind; sometimes the tank was destroyed; in other trials it remained undamaged. At the moment the weapon system fired, one or two lights (according to the number of projectiles that were used) went on in the lower part of the computer screen. The color of the light indicated the kind of projectile. Four lights, corresponding to four projectiles, were used in the experiment: red, yellow, blue, and green.

Conceptually, each light can be considered as a separate cue and the explosion of the tank can be regarded as the outcome. At the end of the experiment, participants were asked to make a judgment about the relationship between each cue (i.e. the light appearing at the bottom of the screen) and the presence of the outcome (i.e., the tank being destroyed). Participants were requested to judge the efficacy of each 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).

The experiment could be conceptually divided in three phases, each phase comprising 15 trials. In every trial of the first phase two projectiles were contemporaneously fired; this phase was indicated by $A, B @ O$. In each trial of the second phase a single projectile was fired that participants of the experimental BB (backward blocking) group had already seen in the first phase ($A @ O$) while it was new for participants of the CC (compound control) group ($C @ O$). In the trials of the last phase participants of both groups saw a series of trials in which a new projectile was fired ($D @ O$). Projectiles A and C had equally high causal power, while B and D were considered as having an equally lower one. The compound A, B had the causal power indicated by the power PC theory. Table 1 summarizes the experiment by reporting the conditions for the two groups, the probability of obtaining the outcome in each phase, and the ratio between the trials with the outcome and the number of trials for each phase.

Table 1: Experimental conditions

BB Group		CC group	
$AB @ O$	0.80 [12/15]	$AB @ O$	0.80 [12/15]
$A @ O$	0.67 [10/15]	$C @ O$	0.67 [10/15]
$D @ O$	0.40 [6/15]	$D @ O$	0.40 [6/15]

According to what has been discussed in the previous section, the power PC theory allows making the following predictions:

1. **Independence from compound:** Judgments about cue B given by participants in the BB group should

be identical to judgments provided by participants in the CC group about cue D : $j_{BB}(B) = j_{CC}(D)$. According to the theory, judgments about the causal power of B should be independent from cue A to which it was paired, and should be identical to those given to cue D having the same causal power.

2. **Irrelevance of compound:** Judgments about cue A given by participants in the BB group should be identical to judgments provided by participants in the CC group about cue C : $j_{BB}(A) = j_{CC}(C)$. Because A and C have the same causal power they should obtain identical judgments, the fact that A had been previously experienced also as a compound stimulus being irrelevant.
3. **Equalization to compound:** Judgments about cue A given by participants in the BB group should be lower than judgments about the same cue provided by participants in the CC group: $j_{BB}(A) < j_{CC}(A)$. For participants of the latter group, the causal power of A should be equal to the causal power of the compound, while participants in the BB group should base their judgments on the trials of the second phase.

Method

Participants Sixty college students (22 males and 38 females) aged between 20 and 26 years (median = 20) participated to the experiment. The data of one participant were missed due to a computer failure.

Stimuli The experiment was performed on a Dell PC computer equipped with a 15" LCD flat screen (with 1024 x 768 pixel resolution) and headphones. A custom-made program written in Java 1.4² was utilized to present the stimuli and to record the participants' judgments. During the presentation trials, the picture of a T-44 soviet 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 4600 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) located at the bottom of the screen, went on. The LEDs could display four different colors: red, blue yellow, and green.

The tank was always hit, and 1000 ms after the LED were brightened, its color changed for a short time (300 ms) to white to simulate the projectile impact. In the

² Copy of the program could be obtained by sending an e-mail to the authors.

trials in which the tank was destroyed, an explosion sound was heard, and the tank was covered by a dust cloud that, after it was 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 10s; after that, with a shutter effect, the view finder was closed and opened again, and a new trial began.

Once in a while, instead of a presentation trial, participants saw a “control” screen with the four LEDs placed at the vertices of an imaginary rectangle positioned at the center of the screen, each LED associated with two radio buttons labeled “Yes” and “No”, respectively. They 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 “not assigned”. 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.

Procedure At the beginning, participants read an instructions sheet, written in Italian, that explained the task. After that, they saw four practice random trials in which a white projectile was fired. The tank was destroyed in two trials, and in the remaining two was left undamaged. At this point the experiment could start.

For each participant, the four colors were randomly assigned to the four projectiles. Participants were exposed to 15 $A, B @ O$ trials in 12 of which the tank was destroyed (hence $P(O | A, B) = 0.80$). After that, they saw a group of 15 trials constituted by $A @ O$ episodes for participants in the BB group and $C @ O$ ones for the CC group. In this case the tank exploded 10 times ($P(O | A) = P(O | C) = 0.67$). Moreover participants saw other 15 ($D @ O$) trials (with 6 explosions: $P(O | D) = 0.40$), that were randomly interleaved within the preceding ones.

To ensure that participants paid attention to the presentation trials, during the experiment five “control” screen 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.

Finally, at the end of the presentation trials participants were asked to provide their judgment about the efficacy of the projectiles.

Results and Discussion

Table 2 summarizes the judgments provided by participants in the various experimental conditions.

As preliminary observation, it is important to note that we were able to replicate the phenomenon of backward

blocking previously found, among others, by Chapman, (1991), Dickinson & Burke, (1996), and Shanks, (1985).

Table 2: Efficacy judgments

	A	B	C	D
BB group	76.9	50.3		50.8
CC group	62.3	60.9	71.3	51.2

The judgments provided by participants in the BB group on cue B were in fact lower ($t(57)=1.64$ $p=.05$) than that provided by participants in the CC group on the same cue. The replication of this well known phenomenon supports the efficacy of the experimental paradigm we adopted to investigate causal judgments.

As far as the first prediction, i.e., independence from compound, is concerned, we were not able to find practically any difference between the judgments given by BB group participants on cue B , and the judgments provided by participants of the CC group on cue D . The first prediction derived from Cheng’s power PC theory seems essentially confirmed, and this corroborates the theory. However, within the limited power of the experiment, ($1-b = 0.60$) we were not allowed to accept the null hypothesis. Anyway, we regard the result as surprising: cue B was never experienced alone, and it was not at all obvious that participants could be able to derive its causal power from the compound stimulus.

The irrelevance from compound prediction, i.e., that the BB group judgments on cue A should be similar to CC group judgments on cue C was not confirmed. In fact, A obtained a higher ($t(57)=1.72$, $p<.05$) rating than C .

Finally, the equalization to compound prediction was disconfirmed by our experimental results, too: the judgment about cue A by the BB group was higher ($t(57)=3.76$, $p=.0002$) than that given by the CC group.

In summary, two out of three predictions from Cheng’s power PC theory about the role of compound cues were falsified by our experiment. It could be interesting to speculate why.

One possible explanation for the effect of the compound on A is that participants make an a posteriori estimation of the frequency of cases in which O is present and absent, respectively. There is evidence for the hypothesis that people have an automatic capability of encoding frequencies of occurrence of the events they attend to (Hasher & Zacks, 1984). It is reasonable to suppose that such an estimation may be influenced by the greater occurrence of A cases associated with a positive outcome O in the compound. In effect, the causal power attributed to cue A by the BB group (76.8) lies midway between that attributed to C by the CC (71.3) and the real causal power of the compound cue (80.0); this result is consistent with participants making use of frequencies of occurrence of cue A both alone and in the compound.

It is interesting to note that the associative models do not make consistent predictions about the role of the

compound. The R&W model, in particular, can only account for results going in the opposite direction, that is $j_{BB}(A) < j_{CC}(C)$. In our design the R&W makes at the asymptotic level the same predictions of the power PC: i.e., the *A* and *C* judgments should be equal. Before reaching the asymptote, on the other hand, the perceived causal power of *A* in BB participants should be intermediate between the asymptotic one and that deriving from previous compound trials. But, according to the R&W, the causal power of the compound should be equally divided between *A* and *B*. As a consequence, the perceived causal power for each of them should be inferior to the asymptotic power, leading to judgments for *A* lower than those given to *C*.

It should be stressed, moreover, that participants, when required to assign a value to *A* and *B* experienced only in compound form, gave judgments that were significantly lower than the causal power of the compound, and quite similar to the estimates provided by the R&W. A possible explanation for this result is that, being unable to disambiguate the causal power or the cues by simple contingency, participant relied to an associative estimate of causal power, showing that such information was available to them and utilizable on demand.

An alternative explanation is that participants would have assigned cues identical causal power, assuming a value capable of producing the real causal power of the compound. Such a value can be estimated by posing $P(O | A) = P(O | B) = p$ in our third assumption's formula, resulting in the second degree equation:

$$2p - p^2 = P(O | A, B) = 0.80$$

The only valid solution is $p = 1 - \sqrt{1 - 0.80} = 0.55$. This value is close to the estimates given by participants to *A* and *B* (i.e., 62,3 and 60,9, respectively).

Conclusions

The results of our experiment suggest the existence of cognitive processes involved in causal judgment that lie outside the scope of the power PC theory. Associative and probabilistic explanations are in fact not only formulated at different levels of Marr's (1982) taxonomy (algorithmic the former, computational the latter) but address also different phenomena. A complete understanding of causal cognition needs probably both, integrated into a unified account. Human causal judgment, as shown by our results and results obtained by Lober & Shanks (2000), seems in particular to be influenced by underlying memory processes that could only be captured by associative models. Probabilistic models, on the other hand, seem capable of explaining some subtle inferences participants are able to draw, as in the case of independence from compound. Anyway, people seem capable of performing a more complex behavior than that supposed by traditional models.

References

- Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgements. *Memory & Cognition*, 23, 510-524.
- Buehner, M. J., & Cheng, P. W. (1997). Causal induction: The power PC theory vs. the Rescorla-Wagner model. In M.G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 55-60). Mahwah, NJ: LEA.
- 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*, 18, 537-545.
- 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.
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information. The case of frequency of occurrence. *American Psychologist*, 39, 1372-1388.
- Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review*, 107, 195-212.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.
- 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 judgments. *Quarterly Journal of Experimental Psychology* 37B, 1-21.
- Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge, UK: Cambridge University Press.
- Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science*, 7, 337-342.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 174-188.