The Impact of the Format of Covariation Information on Causal Inferences

Teresa Payton and Frédéric Vallée-Tourangeau
Psychology Research Unit, Kingston University
Kingston-upon-Thames UNITED KINGDOM KT1 2EE
t.payton / f.vallee-tourangeau@kingston.ac.uk

Abstract

Representational effects in a fictitious virus-disease causal induction task were examined in three studies. In all three studies, six different judgment conditions were created by crossing two levels of virus-disease covariation (0, .5) with three levels of disease base rate (.25, .5, .75). In Study 1, the covariation information was presented as four propositions summarising the frequencies of the four patient types, namely patients with or without the virus who either did have or did not have the disease. In Study 2 the same information was presented in a 2 x 2 table with the cell frequencies represented iconistically (the presence/absence of virus/disease was shown as schematic faces that varied in expression and colour). In Study 3 the covariation information was presented in terms of a branching tree with the two main branches representing the frequencies of patients with and without the disease from which sprouted smaller branches showing the frequency of those with and without the virus. Causal judgments were poorest in Study 1, reflected significantly improved covariation discrimination in Study 2, but were most normative in Study 3. These results signal the presence of important representational effects in causal induction tasks.

Introduction

In formulating a judgment of the causal link between a candidate cause and a target effect, information about the frequency of the pairing of four types of events should be considered. These four types of events refer to the pairings of (a) the cause and the effect, (b) the cause with the absence of the effect, (c) the effect in the absence of the cause, and (d) the absence of both the effect and the cause. These event frequencies are often represented in a 2 x 2 table with the columns referring to the presence or absence of the effect and the rows the presence or absence of the cause (see Fig. 1). The contrast between the probability of the effect in the presence of the cause, P(E|C), and the probability of the effect in the absence of the cause, P(E|~C) is taken as a measure of the covariation between the candidate cause and the effect (also referred to as ΔP).

Early research on reasoners' appreciation of covariation in formulating causal judgments examined how the manner with which the covariation information was presented influenced judgments. For example, in Ward and Jenkins (1965), participants were asked to gauge the relationship between seeding clouds and the occurrence of rain in different regions. The instructions made clear that confirming evidence (i.e., rain) was influenced by climatic variations between the regions as well as (possibly) by the seeding. Information was presented either on-line, where the results of individual trials (seeding/no seeding, leading to rain or no rain) were presented one trial at a time, or the information was presented off-line, that is as static frequency summaries (participants received frequency summaries of the days clouds were and were not seeded, and on how many days in each case it rained and did not rain). Only 17% of the participants receiving the covariation information one trial at a time appeared to formulate judgments in line with the Δ*P* rule compared to 75% of the participants who received summary information.

	E	~E
C	*a*	*b*
~C	*c*	*d*

$$\Delta P = \frac{a}{a+b} - \frac{c}{c+d} = P(E \mid C) - P(E \mid \sim C)$$

Figure 1: A 2x2 contingency table. *a, b,c, d* are cell frequencies; C = candidate cause, E = target effect.

Based on the information available from a 2x2 table, Wasserman, Dorner and Kao (1990) explored the relationship between the information participants deemed important in order to formulate a causal judgment and the information they actually used. Using a drug/disease scenario, their participants either decided which of the four categories of information would be necessary for a causal judgement or rated the importance of each category of information. Cell importance was ranked as cell *a* > cell *b* > cell *c* > cell *d* in both conditions. In a second experiment, using the same causal scenario, problems were structured in quartets of pairs, giving the opportunity to vary the numerical content of one cell while holding constant the contents of the other three. This allowed Wasserman et al. to measure the impact that increasing the frequency in a single cell had on participants' causal judgements. The numerical information from a 2 x 2 table for each problem was presented as a set of four statements. For each problem participants ranked the value of the drug for treatment of the disease on a scale of -10 (drug worsens disease) to +10 (drug helps cure). The importance ratings of each cell

corresponded to those in Experiment 1 (viz. cell a > cell b > cell c > cell d). Wasserman et al. estimated that 21% of the participants used a ΔP strategy to generate their causal judgments, while 50% based their judgments on the comparison between cell *a* and cell *b*.

Improvements in research methodology using on-line causal judgment tasks unveiled a much greater sensitivity to covariation information than was originally suggested by Ward and Jenkins (1965, e.g., Baker, Berbrier, & Vallée-Tourangeau, 1989; Dickinson, Shanks, & Evenden, 1984). Subsequent research (e.g., Vallée-Tourangeau, Murphy, Drew, & Baker, 1998; Murphy, Vallée-Tourangeau, Msetfi, & Baker, 2005) has documented two important features of causal judgments using an on-line procedure: (i) reasoners can make relatively subtle discrimination between different levels of covariation, but that (ii) causal judgements appeared to be influenced by the overall incidence, or base rate, of the effect, *independent* of the actual level of covariation. In other words, the higher the base rate of the effect, the higher the estimate of the importance of the candidate cause (e.g., Vallée-Tourangeau, Hollingsworth & Murphy, 1998).

The dual influence of covariation and the base rate of the effect on cause-effect judgments is illustrated in Vallée-Tourangeau, Murphy, Drew, and Baker (1998, Experiment 1). In that experiment, participants were invited to estimate different types of fictitious virus-disease relationships by sampling, for each type, 40 'patients' who were either infected with the candidate virus or not and who suffered from the target disease or not. The covariation data was presented one patient at a time in a simulated medical diagnosis task. Participants were presented with six different virus-disease relationships reflecting the factorial combination of two levels of virus-disease covariation (0, .5) and three levels of disease base rate (.25, .50, .75). Causal judgments were significantly influenced by both factors. That is, (i) participants judged more positively the positive contingency relationships than the noncontingent ones, and (ii) they judged the virus-disease relationships more positively the higher the base rate of the disease, *independent* of the actual virus-disease covariation.

Note that Vallée-Tourangeau et al. (1998) observed the impact of the effect base rate using an on-line presentation procedure, that is one in which reasoners experienced the event pairings in 'real' time. Such an on-line procedure may engage different information-processing mechanisms that would otherwise be at play were the information presented in terms of event-pairing summaries. For example, the continuous presentation of event pairings may engage associative learning mechanisms that build up the strength of the connection between a candidate cause and the target effect over time, and causal judgments may be a reflection of such associative strength. In fact, associative learning models such as Pearce's (1987) stimulus generalisation model, predict that, for any given level of cause-effect covariation, stronger associative strength will accrue to a candidate cause when it is paired with an effect with a high rather than a low base rate (see Vallée-Tourangeau et al., 1998).

The Present Studies

The purpose of the studies reported here was two fold. The first was to investigate the impact of the effect base rate on causal judgments when the covariation information was presented off-line, that is as summaries of the different kinds of cause-effect pairings, as opposed to being experienced in real time. The second was to investigate whether the manner with which this frequency information was presented modulated in any way the influence of the effect base rate on causal judgments. In judgment under uncertainty, the format of the information presented to reasoners make an important difference in determining the degree to which their judgments reflect a more normative appreciation of the evidence (e.g., Gigerenzer & Hoffrage, 1995). Training participants in different presentation methods (e.g., decision trees) has also been shown to improve performance (Sedlmeier, 2002).

The same causal judgement task was used in each of three studies. The studies differed only in the format of the covariation information. In Study 1 the covariation information was presented in a series of simple statements, while in Studies 2 and 3 the covariation information was provided in graphical formats. The statements provided exactly the same information as the graphical formats in order to provide the same 'space of possibilities' (cf. Stenning & Oberlander, 1995) as the graphical representations, thereby eliminating one of the differences between linguistic and graphical representations.

The graphical formats incorporated a diagrammatical representation of the problem with the information presented as numbers and annotations. Numbers can be represented externally in different ways (i.e., through different numeration systems) and one such representational property is dimensionality. Numbers that are represented as quantities of some object (e.g., stones) have only one dimension. This can be very efficient with small numbers since the representation is proportional to the numerical value. (However, as numbers get larger the system becomes more unwieldy.) In Study 2 the event pairing frequencies were presented in terms of quantities of symbols (schematic faces specifically) that were framed in a standard 2x2 contingency table. Thus, each of the groups of patients that made up one of the cells of the 2x2 table were represented in terms of groups of schematic faces, each face representing a patient. For example, cell *a* patients (those with the virus and the disease) were represented as grey frowning faces, whereas cell *d* patients (those without the virus and without the disease) as white smiling faces.

The information from a 2 x 2 cell format can also be represented in other displays. One of these is a frequency tree. In probabilistic reasoning tasks, training in decision trees or grids has been found to substantially improve

Table 1: Even pairing frequencies in the six experimental conditions, along with the descriptive probabilities. The highlighted conditions pair zero and positive covariation levels with the same Cause-Effect and Cause-No-Effect frequencies (BR = base rate)

	Zero Covariation			Positive Covariation		
	.25BR	.50BR	.75BR	.25BR	.50BR	.75BR
Event Frequencies						
Cause - Effect	5	10	15	10	15	20
Cause - No Effect	15	10	5	10	5	0
No Cause - Effect	5	10	15	0	5	10
No Cause - No Effect	15	10	5	20	15	10
Descriptive Probabilities						
P(Effect)	0.25	0.50	0.75	0.25	0.50	0.75
P(Effect|Cause)	0.25	0.50	0.75	0.50	0.75	1.00
P(Effect|No Cause)	0.25	0.50	0.75	0.00	0.25	0.50
ΔP	0.00	0.00	0.00	0.50	0.50	0.50

success rates in Bayesian reasoning problems (Sedlmeier, 2002). Thus in Study 3, the covariation information was presented in terms of a hierarchical tree structure wherein the top node represented the entire sample of 40 patients which then branched into two nodes that represented the frequencies of the patients with and without the disease. Finally, each of these nodes branched into two new nodes that corresponded to the presence and absence of the virus.

To sum up, the studies reported here used the factorial design employed in Vallée-Tourangeau et al. (1998), combining two covariation levels with different levels of the effect base rate, producing six different cause-effect relationships. Unlike in Vallée-Tourangeau et al., however, the covariation information was presented off-line. In Study 1, each cause-effect relationship was summarised in terms of four propositions describing the frequency of each of the four types of event pairings. In Study 2, the six relationships were presented as 2x2 tables in which symbols represented the different number and type of patients. In Study 3, the event frequencies that defined the level of cause-effect covariation were presented in terms of hierarchical inverted tree structures.

Method

Design & Procedure

All three studies used a questionnaire that described six different virus/disease relationships. These six different relationships reflected the factorial combination of two levels of covariation (0.0 and 0.5) and three levels of disease base rate (0.25, 0.50 and 0.75). Viruses and diseases were assigned fictitious labels (e.g., OHPD Type B and Nachmose A, respectively). The assignment of labels to conditions were counterbalanced resulting in 36 different versions of the questionnaire in each of the three studies. The frequencies of the four different kinds of virus-disease pairings in each of the six conditions are presented in Table 1.

In Study 1 (Propositions) the cell frequencies that defined the virus-disease covariation were presented as a set of four simple statements (the order of statements was counterbalanced across participants). For example in one version of the questionnaires, the following four propositions described the zero covariation relationship at the high disease base rate of .75:

Five people with virus OHPD Type B absent in their blood did not have the disease.
Five people with virus OHPD Type B present in their blood did not have the disease.
Fifteen people with virus OHDP Type B present in their blood also had the disease.
Fifteen people with virus OHDP Type B absent in their blood had the disease.

In Study 2 (Schematic faces) the same information was presented in a 2 x 2 table with the frequencies represented in terms of schematic faces that coded the presence/absence of virus/disease by colour and expression (see Fig. 2). In Study 3 the information was presented as a tree with the two main

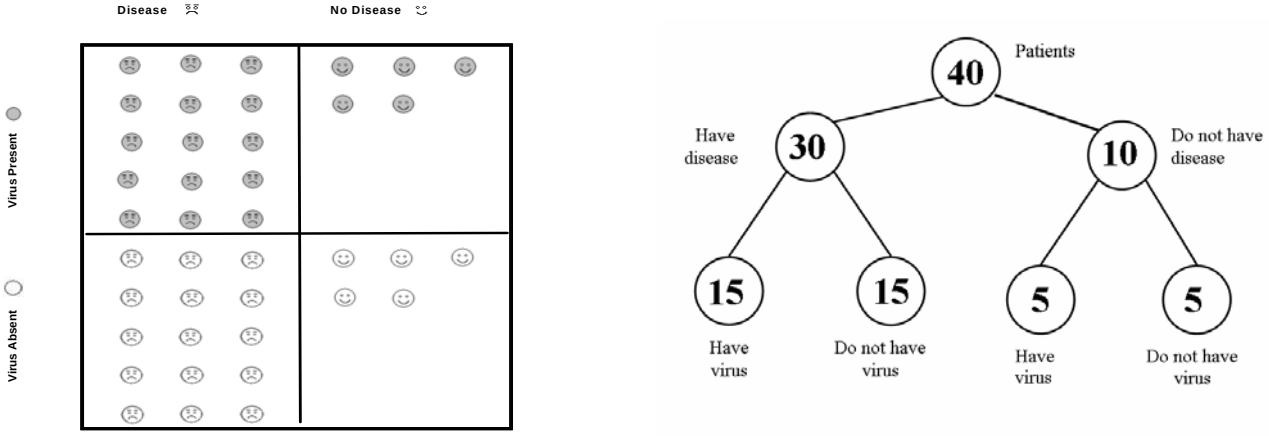

Figure 2: Two different presentations of the same covariation condition (high base rate, zero covariation): Schematic faces in Study 2 (left panel) and tree structure in Study 3 (right panel).

branches representing the frequencies of patients with and without the disease from which branched smaller paths showing the frequency of those with and without the virus (see Fig.2).

Questionnaires in the three studies were identical in all aspects apart from the presentation of the covariation information. Six causal judgements (one for each covariation condition) were required in each questionnaire, with each virus, disease and covariation condition appearing once. The first page of the questionnaire contained the instructions, which were identical for all three studies. Participants were requested to make causal judgements intuitively, making no notes. The scenario was described as being based in a city hospital, where researchers had identified six previously unknown diseases which they suspect have viral origins. To develop successful vaccination treatments it was necessary to identify the virus causing each disease, and blood samples have been collected. For each virus-disease pair, participant were asked to rate the nature of the relationship using a scale ranging from -100 to +100. The rating scale was explained as follows:

> *A **positive** rating indicates that patients who have the virus tend also to have the disease. The more positive the rating, the stronger the relationship. Thus, if you feel the presence of a particular virus always predicts the disease you would give it a rating of 100. If you feel the presence of a particular virus mostly, but not always predicts the disease you might give it a rating of say 85.*
>
> *A **negative** rating means that patients who have the virus tend NOT to have the disease, that is that particular virus somehow affords **immunity** against the disease. The more negative the rating, the greater the immunity. Thus, if you feel the presence of a virus always predicts the disease is **absent** you would give it a rating of -100. If you feel the presence of the virus mostly, but not always predicts the **absence** of the disease you might give it a rating of say -85.*
>
> *A **zero** rating is appropriate when the presence of the virus in the blood does **not** inform you of either the presence **or** absence of the disease.*

Participants

Participants were a sample of 237 undergraduates (Study 1 N = 66, Study 2 N = 86, Study 3 N = 85). Over the three studies, 79% of the participants were female and 21% male (these proportions were approximately constant across the three studies). The mean age of the participants was 23, and did not differ significantly across studies.

Results

In all three studies, overall causal ratings were generally more positive when the virus-disease covariation was positive than when it was zero (see Table 2). Overall ratings also reflected the influence of the base rate of the disease in that ratings of both positive and zero virus-disease covariation levels increased as the prevalence of the disease increased. A two-way repeated measures analysis of variance (ANOVA), with two levels of covariation and three levels of base rate, was performed on the causal ratings for each study. There were significant main effects for both covariation and base rate in all studies, with the interaction being significant only in Study 1 (Propositions; see Table 3).

Table 2: Mean causal ratings (and *standard errors*) in the six conditions for each of the three studies (BR = base rate).

	Zero Covariation			Positive Covariation		
	.25BR	.50BR	.75BR	.25BR	.50BR	.75BR
Propositions	-10.1	-0.9	10.5	-2.9	11.2	48.3
	(4.13)	*(2.69)*	*(3.99)*	*(5.52)*	*(4.31)*	*(5.19)*
Faces	-36.7	10.8	21.5	6.6	43.5	49.1
	(5.34)	*(3.72)*	*(5.32)*	*(5.59)*	*(4.90)*	*(6.11)*
Tree	-8.1	11.7	18.7	23.3	32.3	38.7
	(4.38)	*(3.58)*	*(4.06)*	*(5.84)*	*(4.50)*	*(4.80)*

A second series of analyses were conducted on a subset of four of the six conditions. These conditions were chosen because they permitted control of the probability of the disease given the virus, P(D|V), across the two levels of covariation. This subset contains two sets of 2 contingency tables (one in each covariation condition) with identical top rows and therefore the same probability of the disease given the virus, P(D|V). In the first set P(D|V) = .5 and in the second P(D|V) =.75 (see Table 1). If judgments are driven solely from a consideration of cells *a*, or *a* and *b*, then judgments should not differ across the two levels of covariation.

Table 3: *F* ratios from two-factor repeated-measures analysis of variance including all six conditions (Overall Analysis) and including the four conditions where covariation levels varied over fixed values of P(D|V) (Finer Analysis).

	F ratio		
	Study 1 Propositions	Study 2 Schematic Faces	Study 3 Decision Tree
Overall Analysis			
Contingency (C)	27.83**	41.75**	32.19**
Base Rate (BR)	35.65**	57.78**	9.92**
C x BR	7.37**	1.80	1.48
Finer Analysis			
Contingency (C)	0.02	2.39	6.10*
Base Rate (BR)	8.50**	25.29**	4.01*
C x BR	0.16	11.23**	0.06

* $p < .05$ ** $p < .01$

Figure 3 plots the four means for the conditions where the probability of the disease given the virus, P(D|V), is fixed at either .5 or .75 across the two levels of covariation. A gap

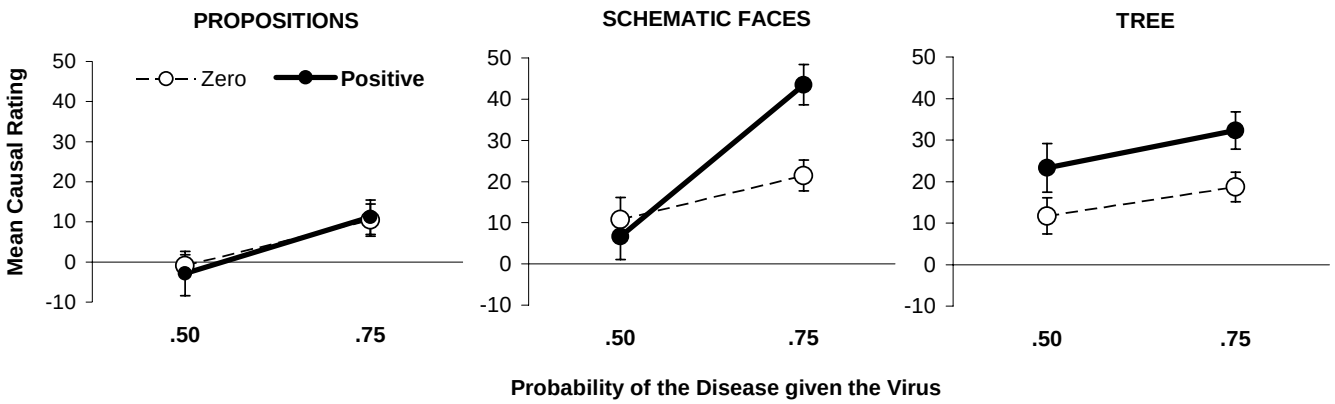

Figure 3: Mean causal ratings in zero and positive covariation conditions when the probability of the disease given the virus is fixed at either .50 or .75, across the three studies.

between parallel lines would suggest a robust discrimination of the two levels of covariation for each of the values of P(D|V). In turn, causal ratings that are more positive with the higher value of P(D|V) indicate the influence of the base rate of the effect in the presence of the cause. As the left panel of Figure 3 indicates, the means for the positive covariation condition completely overlap the means for the zero covariation conditions in Study 1. Thus, participants in Study 1 demonstrated little or no ability to discriminate between positive and zero levels of covariation when cells *a* and *b* were held constant across levels of covariation. However, causal ratings increased as P(D|V) increased, reflecting the influence of the disease base rate on ratings. In Study 2 (middle panel), ratings reflected a discrimination of the positive and zero covariation conditions only when P(D|V) was high, suggesting an interaction between covariation and P(D|V). As in Study 1, ratings were more positive at the higher value of the effect base rate. Finally, ratings in Study 3 (right panel) appeared to reflect a robust covariation discrimination ability across the two levels of P(D|V), although here too, the base rate of effect appeared to have influenced ratings.

A series of two-way repeated measures ANOVAs (2 levels of covariation x 2 levels of P(D|V)) were performed on the data from these four conditions across the three studies (see Table 3). The main effect for P(D|V) was significant in all three. However, confirming impressions, the main effect for covariation was only significant for Study 3 (Tree) and the interaction between covariation and P(D|V) was significant only in Study 2 (Schematic faces, see Fig. 3). A series of mixed ANOVAs were also conducted to determine whether the pattern of means differed significantly between the studies. Using Study as a between-subjects factor, the three way (study x covariation x P(D|V)) interaction was significant in the analysis comparing Study 2 and Study 1, F (1, 150) = 4.80, $p < .05$, and in the analysis comparing Study 3 and Study 2, F (1, 169) = 4.38, $p < .05$. These interactions suggest that covariation discrimination was better using schematic faces to represent the covariation information than propositions, but that in turn the best covariation discrimination was elicited using trees.

Discussion

In the three studies reported here, participants were presented information about the covariation between different virus-disease pairs from which they were asked to make a causal judgment. In all three studies the covariation information was either positive or zero, with the disease base rate manipulated independently. The three studies differed in the format of the information presentation: as propositions summarizing the frequency of each of the four cells of a contingency table (Study 1), as sets of symbols that represented the number and kind of patient types in a 2x2 table (Study 2), and as frequency summaries in a tree-like hierarchical structure (Study 3).

In all three studies causal ratings displayed varying levels of discrimination between the two levels of covariation but in all three, causal ratings at both levels of covariation were significantly affected by disease base rate, that is the higher the base rate, the higher the causal ratings. The significant covariation by base rate interaction in the overall analysis of the means in Study 1 reflected very low discrimination between levels of covariation at the lower base rates and better discrimination at the high .75 base rate level. In contrast, covariation discrimination was observed for all three base rate levels in Studies 2 and 3. These results suggest that judgments of the virus-disease relationships were poorest when the covariation information was provided in terms of four propositions summarising the frequencies of the four patient types.

Results of the finer covariation discrimination analysis when the probability of the disease given the virus, P(D|V), was held constant across covariation levels, suggest that judgments were best, when the covariation data were presented in terms of trees as in Study 3. Covariation discrimination was absent in Study 1; that is, causal ratings when the covariation was positive were indistinguishable as

when the covariation was zero when P(D|V) was fixed. Covariation discrimination was only observed at the high level of P(D|V) in Study 2. In turn, the level of P(D|V) strongly influenced causal ratings in Studies 1 and 2, but less so in Study 3, a finding which again reinforces the observation that judgments were most accurate with the tree-like presentation of the covariation information.

Cross study analyses confirmed these impressions. Using causal ratings from Studies 1 and 2, a mixed ANOVA with study as a between-groups factor and covariation and P(D|V) as within-participants factors, the three-way interaction was significant confirming that genuine covariation discrimination in Study 2 occurred only at the highest level P(D|V). In a similar analysis using ratings from Studies 2 and 3, the three-way interaction was also significant, which in turn confirms that covariation discrimination was superior in Study 3 and that the base rate of the effect exerted the weakest influence on causal ratings in Study 3. These results appear to indicate that the nature of causal judgements can be significantly influenced by the way covariation information is presented.

Taken together, the results of the three studies suggest the presence of substantial representational effect in causal inference tasks. The nature of the covariation data representation influences the degree to which causal ratings reflect the actual degree of covariation. The event frequencies in these three studies were not designed to tease apart the probative value attributed to each cell of a 2x2 table (cf. Wasserman et al., 1990). However, analyses of the causal ratings across covariation levels when the probability of the disease given the virus, P(D|V) was fixed (see Fig. 3) indicated the weighting of the information from the different cells was significantly influenced by the format of the covariation data. The fact that, in Study 1, participants did not discriminate between the two levels of covariation when P(D|V) was held constant (i.e., the overlapping lines in the left panel of Figure 3), implicates a substantial weighting asymmetry in favour of cells *a* and *b*. However, because the values of cells *a* and *b* were not held constant across the two values of P(D|V) in Figure 3. it is impossible to determine whether a consideration of cell *a* alone or cells *a* and *b* actually anchored the causal ratings in Study 1.

Be that as it may, it is clear that the cell weighting asymmetry observed in Study 1 was attenuated in Study 2 since, at least at the high level of P(D|V), participants rated the positive covariation as more positive than the zero covariation even if cells *a* and *b* were held constant across covariation levels, $t(85) = 2.91$, $p < .005$. Furthermore, as the right panel of Figure 3 attests, ratings in Study 3 reflected a robust discrimination of the levels of covariation at both values of P(D|V), when cell *a* and *b* frequencies were held constant. The causal ratings in Study 3 suggest a considerably better balanced consideration of the information provided by the four categories of evidence that define the level of covariation. The results from these three studies provide strong evidence that the format of the covariation data significantly impacted the manner with which information from the four cells of the table was evaluated and integrated to form a judgment.

Everyday reasoners are confronted with many problems that involve static covariation information (e.g., regular exhortation in the media to eat more or less of certain foods to reduce probability of disease based on new evidence). Further research to identify the most effective methods of presenting covariation information could result in the development of cognitive tools that could enhance the layperson's comprehension of probabilistic relationships.

References

Baker, A.G., Berbrier, M.W., & Vallée-Tourangeau (1989). Judgements of a 2x2 contingency table: Sequential processing and the learning curve. *Quarterly Journal of Experimental Psychology*, *41B,* 65-97.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology, 36A*, 29-50.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without Instruction: Frequency formats. *Psychological Review*, *102*, 684-704.

Murphy, R. A., Vallée-Tourangeau, F., Msetfi, R., & Baker, A. G. (2005). Signal-outcome contingency, contiguity and the depressive realism effect. In A. Wills (Ed.), *New directions in human associative learning* (pp. 193-219). Hove, UK: Lawrence Erlbaum Associates, Publishers.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review, 94*, 61–73

Sedlmeier, P. (2002). Improving statistical reasoning by using the right representational format. *Presented at The Sixth International Conference on Teaching Statistics,* Cape Town, South Africa.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. *Cognitive Science*, *19*, 97-140.

Vallée-Tourangeau, F., Hollingsworth, L., & Murphy, R.A. (1998). 'Attentional Bias' in correlation judgements? Smedslund (1963) revisited. *Scandinavian Journal of Psychology*, *39*, 221-233.

Vallée-Tourangeau, F., Murphy, R. A., Drew, S., & Baker, A. G. (1998) Judging the importance of constant and variable candidate causes: A test of the power PC theory. *Quarterly Journal of Experimental Psychology*, *51A*, 65-84

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 19*, 231-241.

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*, 509-521.