

Causal and Predictive Learning Based on Normative Contextualization: The Relevance Relativization Model

Manuel-Miguel Ramos-Álvarez^{1,*} and Andrés Catena²

¹*Departamento de Psicología, Universidad de Jaén, Spain*

²*Departamento de Psicología Experimental, Universidad de Granada, Spain*

Abstract: We present a model aimed at accounting for learning of predictive and causal relationships involving stimulus compounds, by means of a mechanism based on a normative-methodological analysis of causality that goes beyond the traditional associative/rule-based controversy. According to the model, causal learning is attained by computing the validity of each stimulus in a given learning situation. The situation is determined by the assumptions, objectives, and aims held by the learner or demanded by the learning context. Hence, validity computation depends on task demands: causal, predictive, or diagnostic according to a general principle of normative contextualization that allows learners to adapt a between-cues competition principle in a flexible way. Validity is computed using the Relevance Relativization mechanism, a linear model, based on the balance between the probability of stimulus combinations and the probability of each cue. Thus, cue interactions occur mainly when the combination of stimuli shows predictive changes in relation to the same cues considered individually. This model makes novel predictions concerning variations of the competition principle as a function of the type of procedure, including blocking, simultaneous blocking, and relative validity. In addition, our model also integrates top-down and bottom-up processing levels, including individuals' assumptions or previous beliefs.

Keywords: Causal and predictive learning, cue competition effects, predictive validity computation, theory-driven processing.

INTRODUCTION

Learning relationships between events is the basis of many other abilities, such as predictive learning or causal inference in inductive reasoning [1-3]. It is important, for example, to detect danger signals and to find a cause for depression. This depends on our ability to grasp the relationships between the different events in the situation: cue (predictor, cause, or antecedent) and outcome (criterion, effect, or consequent). Hence, the main issue here is to be able to estimate the predictive or causal validity of the cue [1, 4-8].

One common account of causal learning states that the validity of each cue is computed conditionally, and that the estimated validity of any particular cue depends on the validity of the other cues present in the situation. This process is competitive as the validity of each cue is computed controlling for other potential causal cues [9, 10]. The idea that good predictors or causes are weighted more strongly than poor ones is at the core of the relative validity principle, selective learning in associative theories, and discounting in inductive reasoning [8] [11, for a review].

Why should causal learning in complex situations always be solved through conditional computations? Validity can also be computed unconditionally, without discounting the influence of other alternative antecedent events.

Spellman [12] compared conditional and unconditional strategies, and found more support for the conditional one, which reinforces an explanation based on the competition mechanism. Since then, conditional computation mechanisms have been placed at the core of accounts of learning when two or more potential causal agents are present [13]. However, some recent data indicate that when complex potential causal events are involved causal learning can be more flexible, which means that the conditional computational strategy is only one of the possible strategies people can use for solving a learning problem. In the next paragraphs we will further develop this idea.

Firstly, it has been suggested that interaction between cues may emerge, not during learning but rather when a causal response is produced. In other words, conditional computation does not necessarily take place during the coding of relevant information, but when the accrued information is translated into a response or causal judgment. So, theoretical accounts based on the interactive mechanism will explain some performance results, but not how learning proceeds.

The results normally observed in cue competition tasks can be well understood from this view: during the learning stages, people accumulate covariational information about each cue (A and B) separately, so that A and B are both good predictors of the outcome. However, when people are asked for a causal judgment they reason that one of them seems to be a better predictor and base their judgment on its relative validity. This conclusion is supported by recent research [14-

*Address correspondence to this author at the Departamento de Psicología, Universidad de Jaén, Spain; Tel: +34 953212587; Fax: +34 953211881; E-mail: mramos@ujaen.es

16] and constitutes one of our main arguments in favour of this position.

Secondly, although there are practical and theoretical reasons to assume that, normatively speaking, a conditional computation is better for estimating the predictive validity in situations with more than one cue [5, 17, 18], it does not follow that such a computation is *always correct* (see discussions about Simpson's Paradox in [19]). From a methodological point of view, some authors [20] have stated that relative (conditional) computations and independent (unconditional) ones relate to different questions about data. This implies that there is not one correct strategy.

In fact, the association between factors is not a characteristic or attribute of them by themselves, because the associations depend on the contexts where they appear. Let us consider the allergist task (e.g., [21]) in which the reasoner takes the role of an allergist who is asked to rate the strength of different foods as causes of an allergy. This task could be interpreted in the mental context of a single cause model, in which one particular food is supposed to be causing the allergy, so that conditional computation is normative. However, if the task demands or previous beliefs lead the learner to observe (and rate) the relation or co-occurrence of the different foods with the allergy, the model under consideration would likely be a multiple-cause one. When multiple cues can cause the same outcome, the best strategy is to compute the causal strength of each one separately, but bearing in mind the potential cooperation between them at producing the outcome. In this vein, Matute *et al.* [22] have shown that task demands can determine whether cue competition will be observed. Put simply, the two judgments obey different demands, and both are relevant. Instead of assuming a single strategy, it is reasonable to expect participants to adopt either of the two perspectives, so that conditional or unconditional computations are used in a flexible way.

One good example of the flexible use of the conditional strategy comes from studies that manipulate causal directionality: in these, the temporal structure (or order of presentation of stimuli) is opposite to the causal structure (or task-imposed context) of cues and outcomes. In the predictive condition, the information concerning causally antecedent stimuli is presented first, and information about causally consequent events afterwards. However, when making a diagnosis, the order is the opposite: causally subsequent stimuli are considered first, and then causally antecedent ones. For example, in a medical context [23, 24] a set of illnesses can be presented as potential causes of a given common symptom (predictive directionality, illness→symptom), but also a set of symptoms can diagnose a given illness (diagnostic directionality, symptoms←illness). Independently of the order in which illnesses and symptoms are presented, symptoms are effects, and illnesses causes. Competition between events is customarily observed when these events play the role of causes of a common effect (predictive context). However, it has been claimed that no competition occurs when events are seen as effects of a common cause (diagnostic context). Thus, causal directionality appears to determine whether it is sensible to apply the competitive strategy [25]. Although this asymmetric effect is not undisputed [26], many authors accept that previous relevant causal knowledge can affect the processing of subsequent causal information.

We also assume that acquired knowledge guides predictive and causal learning, that is to say, learning is not only data-driven (driven by evidence) but also theory-driven [27, 28], so that the relativization mechanism is driven by task demands and also by people's aims, beliefs and knowledge about the situation. This general principle has been somewhat supported in the causal learning literature [29], but, to the best of our knowledge, never in the predictive learning literature. The main aim of the model we develop here is to set forth the conditions under which relativization occurs and when it does not, including those related to the emergence of cue competition at the time of judgment.

Finally, the different procedures employed in competition research (blocking, overshadowing, conditioned inhibition, superconditioning and relative validity) might not be equivalent. There are reasons to consider that different procedures could lead to the use of relative strategies in some cases, and independent ones in others. For instance, a relative validity procedure could activate a relative computation strategy, because the simultaneous comparison between diverse types of experience – compound and compounding stimulus – is made feasible, thus the apparent correspondence with conditional computations. However, standard blocking favours an independent computation strategy for each predictor, as it fosters a segregation of the cues across the two phases of the procedure. Our model predicts differences in the use of relativization, depending on the features of the learning experience.

Our theoretical framework for the competition mechanism should clearly delimit assumptions (or beliefs) and objectives relevant for predictions and causality judgments, making explicit the conditions that govern the application of relativization. We assume that computations will depend on several principles: 1) Individuals can use different types of computation, depending on their prior suppositions, the type of context, their aims, and their knowledge of the situation; 2) different experimental procedures lead to different learning experiences that will also affect how and when relativization is applied, and (3) the model should explain transfer between learning (or acquisition of information about relations between events) and how performance based on such learning occurs. How can the cognitive system display responses that reflect the competitive action mechanism and at the same time learn about cues in an isolated way? Answering this question is, in essence, our aim in proposing a relevance relativization model.

NORMATIVE ANALYSIS FOR THE COMPUTATION OF PREDICTIVE VALIDITY

In recent decades predictive-causal learning tasks have been designed with a normative framework in mind. In other words, contemporary research tends to use experimental designs (and predictions) based on a high degree of formalization. This agenda allows us to obtain magnitude estimations of the relationship between events with probability calculus. This approach comes from the Focal Set Contrast model developed by Cheng and Novick [5, 17, 30], where validity is derived from δP ,

$$\delta P = \underbrace{P(Y/i)}_{P_Y} - \underbrace{P(Y/\bar{i})}_{P_N} = \frac{a}{a+b} - \frac{c}{c+d} \quad (1)$$

δP is computed as the difference between two conditional probabilities: the probability of an outcome (Y) given the predictor i , P_Y , or *predictive ability of the cue*, minus the probability of the outcome given the absence of the predictor, P_N , or *predictability of the consequence* [31]. The estimation of these probabilities is based on the frequencies in the contingency matrix: the number of cases where both predictor and criterion appear (a), cases in which only the predictor is present (b), cases with only the result (c), and cases without either the predictor or the consequence (d). Hereafter, we will refer to the canonical sequence order a, b, c, and d of the cells in the contingency table.

The higher the discrepancy between P_Y and P_N , the stronger the relation between cue and outcome will be. For example, let us suppose that in a sample of 60 cases, 40 ate shrimps and 20 did not. Among those that consumed shrimps, 36 developed an allergy. Thus, P_Y is 36/40. Among those 20 that did not eat shrimps, only four developed an allergy, so that P_N is 4/20. Formally, the relation obtained from δP would be .70 ($\delta P = 36/40 - 4/20 = .90 - .20 = .70$), namely, a positive high magnitude.

An adequate generalization to situations with more than one cue requires the use of δP conditioned on the remaining factors. That is to say, the validity magnitude for each predictor has to be determined controlling that of other potential predictors. Thus, we should select circumscribed focal sets of evidence where that control can become effective. For instance, whenever the rule δP is operative, it should be applied to each of the relevant cues, keeping constant the remaining ones at their two possible values, present or absent. Computations are similar to those of δP on Equation 1 but selecting those sets of stimuli relevant to each conditionalization test.

Within this framework, the problem arises when comparing the information coming from the two focal sets, or the simple effects in experimental design theory [32]. For instance, if we were interested in estimating the predictive validity for cue i in the presence of another set of cues, j , according to the conditional perspective we would need to compute the simple effect of i when j is present, and also the effect of i when j is absent. According to the Probabilistic Contrast model of Cheng [17], these computations could lead to different results, which would in turn lead to situations with different causal interpretations, as for instance, the difference between causes and enabling conditions. However, what happens if any of the estimations/computations remains indeterminate? In such cases, psychological consequences are unclear.

From our point of view it would be more practical to carry out computations by using the general linear regression model (for instance [33]). In this approach, the relative predictive validity of each cue is computed as the parameter of the regression equation, namely, the slope of the linear function between the target cue and the criterion, discarding the influence of other factors. This model is connected to the probabilistic learning tradition [34] and the lens model of Egon Brunswik [35]. For two cues the validity of the predictor i , relative to other predictors, j , is computed as:

$$\beta_i = \frac{S_Y \cdot (r_{i,Y} - r_{i,j} \cdot r_{j,Y})}{S_i \cdot (1 - r_{i,j}^2)} \quad (2)$$

Where the slope (β_i) is the product of two factors: (1) the ratio between the outcome standard deviation (S_Y) and the standard deviation of the target cue (S_i); and (2) the relativization factor, where $r_{i,Y}$ is the Pearson product-moment correlation coefficient between predictor i and outcome, $r_{j,Y}$, is that between predictor j and outcome, and $r_{i,j}$ is the correlation between the predictors.

THE RELEVANCE RELATIVIZATION MODEL (BASED ON NORMATIVE CONTEXTUALIZATION)

Some researchers have suggested an information integration model which shares some features with the linear model [36, 37]. Our proposal extends this model to binary variables (for a review of the non-metric variant, see [38]) from the predictive point of view. Our model assumes that the estimation of the validity of the target cue has three components: computation of the raw validities, determination of the relevance of the relativization, and computation of the relative validities. The relativization is only applied to the cues included in the set of relevant events (Ω). Relativization is computed as in linear regression, by fine-tuning the slope. This coefficient rearranges the global validity of a cue adjusting it as a function of two computations: the global validity of the remaining alternative cues relevant for discounting (\tilde{r}_j), and the inter-relation among cues ($\tilde{\phi}_{ij}$). The estimations of all the components from our model appear in Fig. (1) and an example of their applications is illustrated in Fig. (2).

The Computation of the Raw Validities

While people are learning, globally computed validity values would be stored as raw estimations of the relation between these cues and the outcomes. According to Equation 2, this could be obtained from the slope for each cue i , removing all other cues. Specifically, the raw validity of cue i is the product of the (estimated) relative rates of event occurrence, ($\tilde{S}_Y / \tilde{S}_i$), and the global predictive value of the stimulus being updated (\tilde{r}_i). This computation is similar to the unconditional one, in the sense that it is made for each predictor in an independent way.

The estimation of global predictive values for binary variables can be made according to two types of rules, one of a normative kind (Alpha), and another one of a heuristic type (Beta) (see [39], for a discussion of the two strategies). The first is an algebraic combination of conditional probabilities, such as δP or any other of a directional character; the second strategy is a linear one, based on frequency counts, and it would not be correct in the normative sense although it shows a high correlation with normative indices [40].

$$\begin{aligned} \text{Alpha: } \tilde{r}_{i \rightarrow Y} &= \frac{W_a a}{W_a a + W_b b} - \frac{W_c c}{W_c c + W_d d} \\ \text{Beta: } \tilde{r}_{i \rightarrow Y} &= W_0 + \frac{W_a a - W_b b - W_c c + W_d d}{W_a a + W_b b + W_c c + W_d d} \end{aligned} \quad (3)$$

W_a to W_d are weights reflecting the subjective importance the person attributes to each of the cells in the contin-

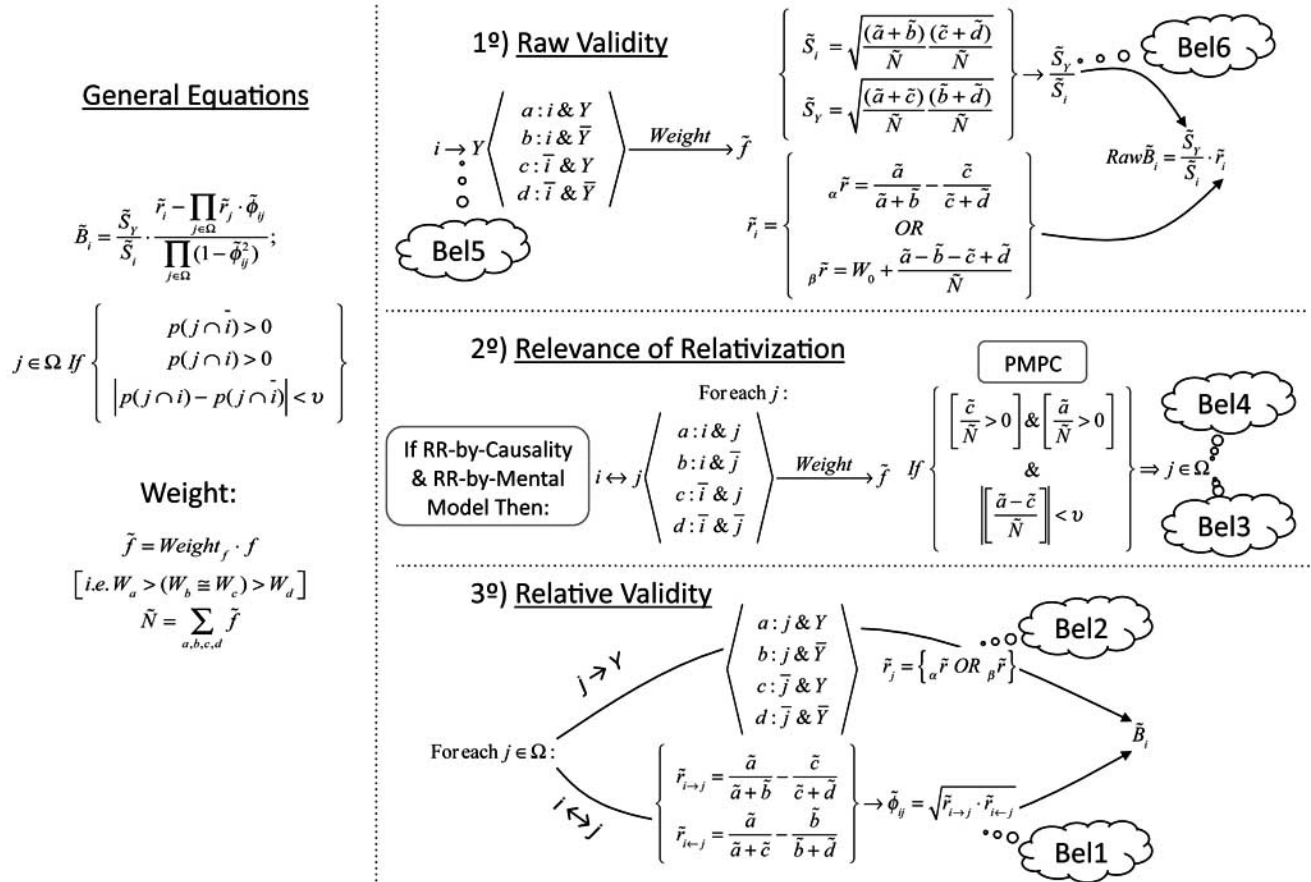


Fig. (1). Relevance Relativization Model (based on normative contextualization). The figure follows the order of the explanation in the text: the general equation, the raw estimation of validity of the target cue *i*, the computations for relevance of relativization to form the set Ω from the three Principles: Probabilistic Mechanism for Predictive Changes (PMPC), RR-by-Causality, & RR-by-Mental Model; the calculations of relative discounting of stimuli in such a set –predictive validity of each alternative cue and their relationship with the target. We have incorporated the effects of theory-driven processing on the model, indicating the place where assumptions or beliefs would have an effect (Bel1 to Bel6).

gency matrix [41]. W_0 represents an initial bias to be used in an anchoring-and-adjustment strategy [36, 42].

For illustration purposes (see Section 1, Figs. (1 & 2), taking into account that the target stimulus in Fig. (1) is Chicken –C– in Fig. 2), let us suppose that prediction of an allergy is made on the basis of two types of food: shrimps and chicken. Eating shrimps is a valid predictor of the allergy, as its raw validity is high, .70. However, eating chicken is a poor predictor as its raw validity is 0.35. The computations for chicken are as follows. In our example, all the events have an equivalent rate of appearance, a standard deviation of .47¹. The product of the relative rate of the event ($\tilde{S}_y / \tilde{S}_i = .47/.47$ in the example) multiplied by the global predictive value of the stimulus being updated ($\alpha \tilde{r}_i = .35$) yields the raw validity value we are looking for, $(1)(.35) = .35$. This is also the way in which computations for the other predictor (shrimp) would be carried out.

The Relevance of Relativization

As we mentioned in the introductory section, relativization tends to occur when a response is required, so that during learning, globally computed validity values are stored as raw estimations of the relations between these cues and the consequence ($\text{Raw} \tilde{B}_i$ in our model). When a causal judgment has to be produced, the relativization mechanism will be triggered, leading to the computation of the relative validity (\tilde{B}_i in our model).

Occasionally, the literature shows considerable processing of stimuli that have no control over behaviour [43]. In fact, in animal conditioning research, studies about revitalization of responses have shown that learning did take place about apparently shadowed or blocked stimuli [44, 45]. This has led some researchers to conclude that statistical dependencies among cues are learnt without interaction. However, interaction can occur when a judgment is to be made, so two separate mechanisms have been proposed: an associative one that would regulate the coding and an inductive one which would determine the final response [14]. Our model provides a normative framework for learning-performance transfer,

¹ The variability in binary variables is computed through the product of the probability of each of the two possible values, $P(j)P(\bar{j})$. For instance, if cue *j* appeared in 80 out of 120 trials, then the variance of *j* is: $(80/120)(40/120) = .22$. The square root of this product is the standard deviation of *j* (S_j).

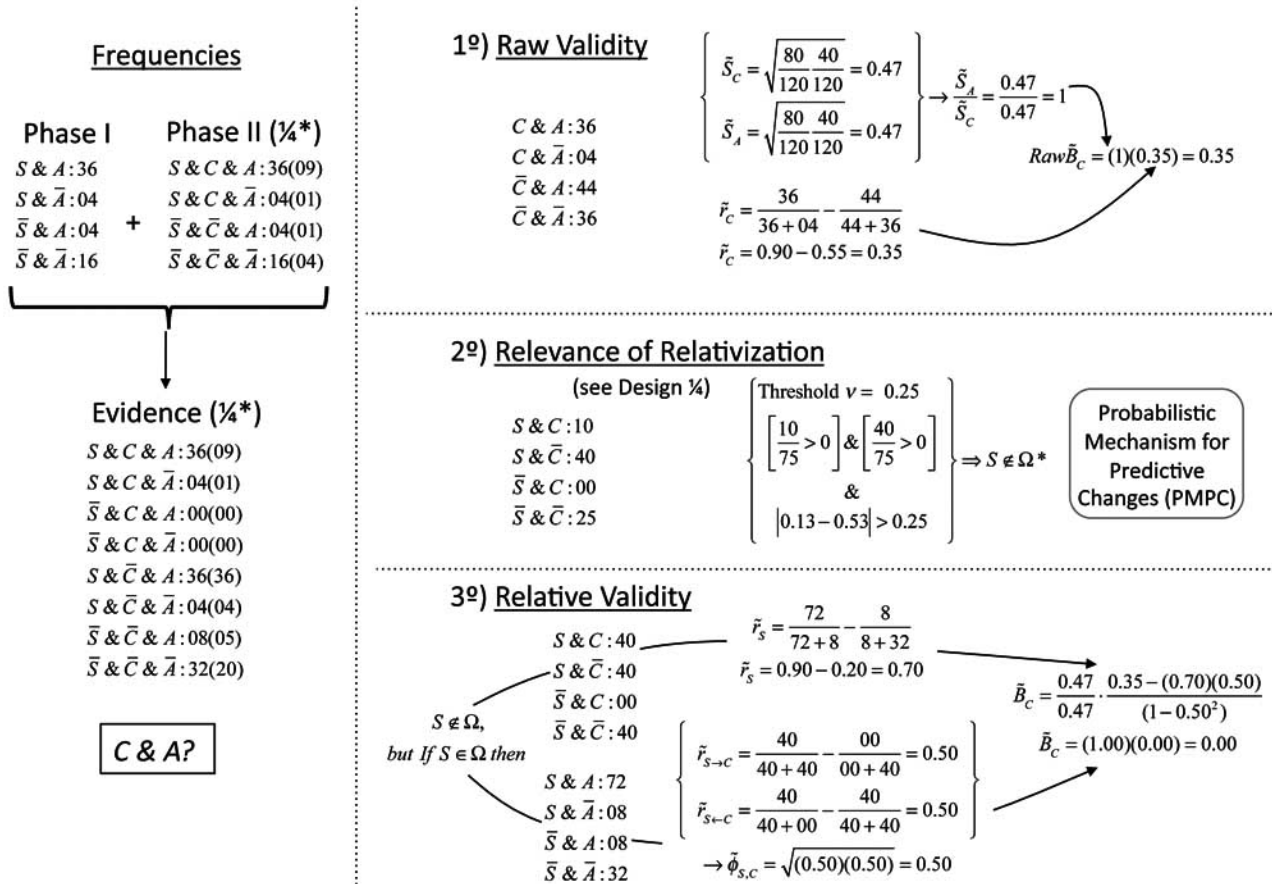


Fig. (2). Computations for the Relevance Relativization Model. The left part shows the frequencies in a probabilistic blocking design, so that the frequencies of the two phases are integrated in the table that appears below the evidence heading. The numbers within charts symbolize what would happen at the beginning of phase two (first quarter). The stimuli are: shrimp (S), chicken (C), and allergy (A). The details of the computation correspond to the equations of Fig. (1). C stands for cue *i*, and A for outcome *Y*.

and, besides, goes one step further as it is able to formally explain such transfer.

On the other hand, having to produce a response does not imply that the relativization principle must be employed. As we pointed out before, there are arguments that support the idea that the cognitive system is flexible, so that the learner could apply the relativization principle or not, depending on the circumstances. As a matter of fact, subjects' performance in learning tasks with stimulus compounds cannot be judged in absolute terms simply depending on whether they match conditioned computations or not. This leads us to a key principle: the relativization mechanism has to be supported by an adequate normative contextualization.

First, our aim in a learning situation, in other words the processing demands imposed by having to produce a predictive or a causal judgment, will determine how the mechanism is employed. As Bindra *et al.* say: "What is of importance adaptively for an individual is the fact that in a stable environment, particular events or conditions usually have some predictive significance—they imply that some other particular events are likely to occur subsequently. Such a relation of unidirectional prediction may exist between two events without a logical, bidirectional relation or a causal, generative connection between them" [46, p. 423]. Learning that there are particular predictive relationships/connections

does not necessarily imply that we learn that the first stimulus is the cause and the second is the effect [13]. Simply put, we learn that both are related and that one of them precedes the other.

It is also true that predictive learning constitutes, together with other factors, part of the causal process. Establishing causality goes one step further, as, it additionally involves the specification of the mechanism responsible for the connection/relationship [17]. Consequently, learning about possible connections, associations or relationships between events is an aspect which is more primary than establishing causality. If the situation is merely predictive, we would learn about each potential causal agent separately, establishing its validity in an independent way ($Raw \bar{B}_i$). Then we would include such estimations in relativization computations (\bar{B}_i) when searching for the most probable cause of the consequent event. That is to say, *relativization would be relevant when establishing causality (hereafter, RR-by-Causality)*. This could explain results such as those found by Matute and her colleagues [16, 47] showing that the type of question (contiguity, causality, predictive) and scenario (instructions and labels) have a modulating role on judgments in cue competition tasks. Pineño *et al.* [16] found competition between two causes when assessed with a causal test

question (i.e., *Is the allergic reaction caused by eating chicken?*) but not when assessed by a predictive test question (i.e., *Is the allergic reaction predicted by eating chicken?*). Moreover, competition was not found between predictors, when assessed by either predictive or causal test questions.

Secondly, different competition paradigms could lead to the use of relative strategies in some cases and independent ones in others. For instance, Wagner's [48] relative validity procedure favours the simultaneous comparison between cues and the compound, and hence favours a conditional computation. However, standard Kamin blocking [49, 50] fosters a segregation of the experience with the compound and the elements through the two phases of the procedure; and hence it favours an unconditional computation. These two paradigms show the two extremes of a continuum, with relative validity at one extreme and blocking at the other, and leave other phenomena (i.e., overshadowing, simultaneous blocking, etc.) between these two extremes. Relativization is only feasible if the procedure includes at least two cues, and these cues are trained together. Otherwise, there would be no reason to learn that these cues have something in common. Consequently, relativization depends on the probability of the conjunction of the cues as compared to the probability of the stimulus.

Formally, in order to compute the validity associated to the target cue i , it would be relevant to relativize with regard to an alternative stimulus j when two conditions are met. First, cue j has to have a minimum probability of appearance in isolation, $p(j \cap \bar{i}) > 0$, and in combination with the target cue, $p(j \cap i) > 0$. And second, when these probabilities are similar: $|p(j \cap i) - p(j \cap \bar{i})| \approx 0$. From now on, we will refer to these computations as the Probabilistic Mechanism for Predictive Changes (PMPC). Fig. (3) illustrates how it operates. If the probability of the union is similar to the probability of the alternative cue in isolation, then the combination may introduce predictive changes (see Situation 1 on Fig. 3) and j could change what was learned about i . However, if the probability of the union is lower than the isolated probability (Fig. 3, Situation 2), then the conjunction cannot improve the prediction in a significant way, so that the alternative cue is not considered for discounting or relativization. This is because j seems to have nothing in common with i . Likewise, if the probability of the conjunction is higher than that of the isolated cues, the latter would not be considered for discounting because the alternative cue is not significant regarding prediction. In this case, there is no perception that they can modulate each other as they appear to be a single stimulus (Situation 3 in Fig. 3).

PMPC would determine that relativization is pertinent below a threshold. The threshold will depend on the complexity of the situation, particularly the number of cue combinations as compared to the number of isolated cues. In the remaining sections we will assume that .25 is an appropriate estimate for PMPC, $|p(j \cap i) - p(j \cap \bar{i})|$, given that the simulations we have carried out with causal learning tasks yield values between .00 and .50 [51].

Returning to the allergist task, Fig. (2) shows that eating chicken has validity between .00 and .35, according to

whether or not relativization is applied. If the task is of a standard blocking type, the PMPC would lead the system towards an unconditional (non relative) strategy, and select a moderate validity of .35. This is because the i cue does not fulfill the probabilistic conditions at the beginning of the blocking phase (i.e., first quarter). Despite the fact that neither probability is null (10/75 or .13 for $p(j \cap i)$ and 40/75 or .53 for $p(j \cap \bar{i})$) the difference between them, .50, is higher than the threshold value, .25. Details are displayed in section 2 of Fig. (2), bearing in mind that frequencies appear within parentheses in the Tables in the top part (1/4*).

The Computation of Relative Validity

Considering raw validity as the starting point, the reasoner could then filter it taking into account the validity of other alternative potential predictors, following a relative perspective which, statistically speaking, closely converges with the computation of conditional contingency. This relativization consists of subtracting from the raw validity the global predictive value of the alternative cue (r_j), weighted by the strength of the correlation between the two predictors. The estimation of the inter-relations among predictors could be obtained by any appropriate algorithm for the estimation of the statistical association, such as the ϕ coefficient² for binary variables [52]:

$$\tilde{\phi}_{ij} = \sqrt{\left(\frac{a}{a+b} - \frac{c}{c+d}\right) \cdot \left(\frac{a}{a+c} - \frac{b}{b+d}\right)} \tag{4}$$

Where a , b , c and d are now the frequencies of the combinations of the two cues i and j (see Fig. 2, section 3). It is easy to show that equation 4 is the geometric mean of the directional coefficients $\tilde{r}_{i \rightarrow j}$ and $\tilde{r}_{j \rightarrow i}$.

The predictive validity of the target cue i can be estimated according to:

$$\tilde{B}_i = \frac{\tilde{S}_y}{\tilde{S}_i} \cdot \frac{\tilde{r}_i - \prod_{j \in \Omega} \tilde{r}_j \cdot \tilde{\phi}_{ij}}{\prod_{j \in \Omega} (1 - \tilde{\phi}_{ij}^2)}; j \in \Omega \text{ If } \left\{ \begin{array}{l} p(j \cap \bar{i}) > 0 \\ p(j \cap i) > 0 \\ |p(j \cap i) - p(j \cap \bar{i})| < v \end{array} \right. \tag{5}$$

This is in essence a regression strategy. First, the global validity of the stimulus being updated (\tilde{r}_i) is estimated, as well as the dispersion of the samples, that is, the rate of outcome appearance (\tilde{S}_y) relative to rate of the target cue (\tilde{S}_i). Then a circumscribed set of potentially competing cues is generated (Ω) which must be also included in the relativization mechanism ($j \in \Omega$). This set would be mainly, but not exclusively, limited by their pertinence or causal relevance [53] together with the background knowledge of the individual. A cue j goes over to this set of alternative stimuli (Ω) when it introduces predictive changes. Causal scenarios tend to induce relativization, whereas non-causal ones (particularly

²The Phi Coefficient is defined as the normalised difference of the two diagonals in the contingency table, $\phi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$. It can be easily shown that this expression is equivalent to the geometric mean of $\tilde{r}_{i \rightarrow j}$ and $\tilde{r}_{j \rightarrow i}$.

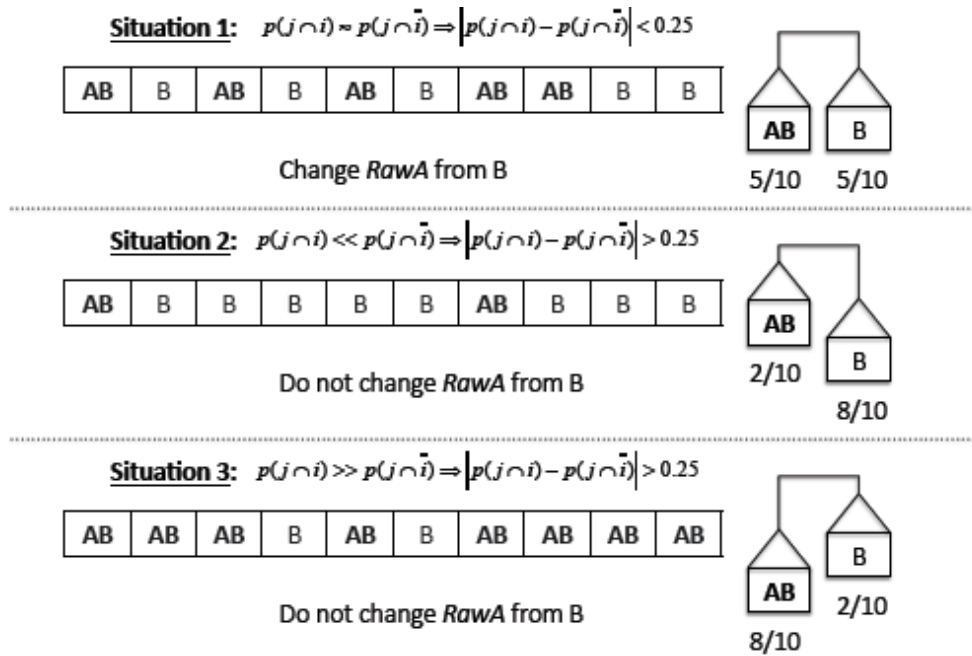


Fig. (3). Illustration of the formal mechanism that governs the relevance of relativization depending on the type of learning experience (PMPC). Cue A is indicated by *i* and cue B by *j*. In situation 1, relativization is pertinent because predictive information is increased when cue A is added. In situations 2 and 3 relativization is not pertinent because the compound is far less frequent than cue B (2) or because cues A and B appear to be too strongly associated (3). The threshold level has been set at 0.25.

of the predictive type) do not. Still, PMPC can alter this situation: for example, in causal scenarios there could be no relativization when the probability of a conjunction does not add changes with respect to the probability of the elements. In a predictive scenario, on the other hand, relativization will not take place during the coding of information, but it could intervene when translating what has been learned into performance.

Good alternative cues modulate the validity of the target cue when they appear simultaneously. In this way we would go from raw towards relative estimation following a simple mechanism with no further assumptions beyond those of any other model proposed in the predictive-causal learning literature, such as the Focal Set Contrast model (see Equation 1). If we focus on the food chicken (*i* in Fig. (1), “C” cue in Fig. 2), we have all the necessary values except the inter-relation, which is estimated as the χ^2 coefficient [52] (see top part of section 3 on Figs. 1 & 2). Thus, following equation 5, we can compute the relativization factor as the ratio $\frac{RawB_i - \tilde{r}_j \tilde{\phi}_j}{1 - \tilde{\phi}_j^2}$.

In our example the relativization factor for chicken would be $(.35 - (.70)(.50)) / (1 - .50^2) = 0$, leading us to a relative validity of .00. In sum, raw validity, .35, and relative validity, .00, for the food chicken differ in this situation.

THE ROLE OF KNOWLEDGE IN RELATIVIZATION

Our model assumes that previously acquired knowledge guides predictive and causal learning; thus, knowledge can determine when relativization is pertinent. To be precise, if the learner imposes a cause-to-effect model (i.e., illnesses \rightarrow symptom) then relativization will be applicable. By contrast, in a diagnostic model of the type effects-

to-cause (symptoms \leftarrow illness), relativization will not apply. This is because people understand that cues can compete if they play the role of causes of a common symptom (when they are antecedent events) but not when they play the role of multiple effects of the same cause (when they are consequent events). The effect of the causal mental model comes before the effect of the directionality observed in the task [27], and we can assume that it also comes before the effect of other features of the causal scenario induced by labels or instructions [54]. Still, the effect of the causal model can be ruled by type of question. When the question matches the causal direction (probability of Effect/Cause: “*Is eating chicken indicative that the allergic reaction is going to occur?*”) competition would take place between stimuli described as causes with a common effect, but not between effects of a common cause. The effect would be the opposite with predictive questions (probability of Cause/Effect: “*Is the allergic reaction indicative that the patient has eaten Chicken?*”). Finally, contiguity questions, with no predictive or causal value, such as “*On the occasions in which the patient has eaten Chicken, did an allergic reaction occur?*”, do not entail competition [22].

Taking all this into account, it can be concluded that relevance relativization guided by a mental model (RR-by-Mental-Model) or by task demands (RR-by-Causality) will jointly overrule the principle that depends on task data features (PMPC).

Knowledge can be made concrete with different generalizability levels [55]. It can be in the form of domain-specific beliefs, but it can also be knowledge about structural properties, with a general character across different domains (i.e., meta-beliefs), including novel contexts. Among some of these properties we can find those of directionality,

variability, or even certain sample features such as its size [4].

Therefore, if knowledge plays a role in the learning process, expanding the task to stimulus compounds would entail thinking that conceptually-driven processes could also be adapted to the principle of predictive competition. Morris and Larrick [29] identified a series of factors which can modulate the effect of the competition principle within the domain of causal attributions in diagnostic normative situations (once an effect, i.e. a symptom, is observed, the task of the individual is to infer out which among several is the cause, i.e. illness, of the observed effect). The processing of diverse events can lead to either of the two results, competition or its absence, depending on a set of assumptions relevant to stimulus compounds concerning a priori likelihoods, independence, sufficiency of integrating elements, and also concerning the number of alternative causes. Depending on these assumptions, sometimes competition can be obtained, but this is not always the case. In our model, prior knowledge is relevant at several levels:

- Regarding the interrelation factor, $\tilde{\phi}_{ij}$, we assert the importance of previous beliefs about the relationship among cues (intra-compound associations within the associative framework, dependence in inductive terms; Bel1 in Fig. 1), but also about the previous validity (\tilde{r}_j) of the remaining cues (a specific belief or the a-priori belief fixed beforehand, Bel2 in Fig. 1).
- Assumptions regarding sufficiency and number of alternative cues, Bel3 and Bel4 (Fig. 1), would have an effect on the potential relativization set (Ω), restricting it, so that these assumptions could be added to the RR-by-Causal principle.
- We also assert the relevance of some other factors, which have no direct connection with stimulus compounds, such as directionality, variability or sample size. For instance, the directionality perceived by participants would lead them to apply a regression algorithm (Equation 2) in one direction or in the opposite direction, similarly to what happens in the General Linear Model [56]. Alternatively, we could also estimate global validity (\tilde{r}_i) by using the direction opposite to δP ; simply switching the role of information type b and c in the contingency matrix (Bel5 in Fig. 1). Variability plays a direct role in the first ratio in Equations 2 and 5. This is the way in which our model is able to incorporate the density of events (Bel6 in Fig. 1). Furthermore, effects such as density of consequences [57] or assumptions about variability of samples are subsumed within the general principle according to which differential weights are attached to different types of experience. If we connect studies which found a mismatch of the type $W_a > (W_b \approx W_c) > W_d$ with associative model proposals [58] which attach more importance to the consequence than to the antecedent event, then we arrive at the inequality $W_a > W_c > W_b > W_d$. If we apply this inequality on the \tilde{S}_y/\tilde{S}_i ratio, it is clear that the numerator would prevail, rather than the denomi-

nator (Fig. 1), hence the effect of density of consequences.

- Finally, we assume conceptual influences in different levels of computation. For instance, the sample rate may cause absolute changes which have nothing to do with the relativization operations, that is to say, they simply produce a bias. However, people's assumptions about stimulus compounds, such as a priori probability, independence between cues, sufficiency of intervening elements and number of alternative events would operate within the computations of relativization.

APPLICATION OF THE RELEVANCE RELATIVIZATION MODEL

The competition principle has been extensively studied using different procedures, which can be classified into three categories. In the best known one, standard blocking, the degree of learning about X in AX+ pairings is attenuated by prior exposure to A+ pairings (A+, AX+); cue A blocks the learning about cue X [49, 50]. This effect has been frequently replicated in the context of predictive-causal learning [11]. The second procedure is similar to blocking but involves comparing responses to different cues simultaneously appearing in the task (e.g., Z vs X in an A+; B-; AX+; BY+; WZ+ design), rather than comparing conditions across different groups of participants [23, 59]. Competition in this second procedure has also been replicated (see [14], for a review, and [60]; for recent studies). The third procedure, relative validity [61], can also be understood as blocking but in this case the different types of trials with single and compound stimuli (A+...AX+) are mixed together in a single undifferentiated stage. Relative validity has also been replicated [62, 63]. However, some variations of these procedures do not generate competition [64], or do so to a diminished degree [21].

From our point of view, validity will be relativized, and thus competition-like effects will occur, whenever there are alternative predictors that involve changes when combined but not when appearing in isolation (PMPC). This disqualifies the inclusion of cues which occur only on rare occasions as compared to a highly frequent target cue. We also discard events which form a configuration with the target. Consequently, we assume that (1) an independent computation strategy will be used when the presence of trials with isolated elements (A+) is higher than that of trials with the compound (AX+), whereas a relative strategy will be used when there is a predominance of compound trials (AX+). This explains how cues may acquire properties in isolation or only in combination. Thus, in a standard blocking procedure (see panel A in Table 1), the strategy would be unconditional, as the density of trials A+ is higher than that of trials of the type AX+ from the beginning of the task, whereas within a relative validity procedure the strategy would be relative, as the density of trials A+ and AX+ is similar from the start.

Simultaneous or within-subjects blocking would be an intermediate case, more similar to validity than to standard blocking. Table 1 displays a direct comparison of these three types of procedure. Frequencies of each type of trial have been adjusted so that they can be directly compared. The

Table 1. Application of the Relevance Relativization Model to the Main Competition Tasks

A. Designs of Three Common Competition Tasks

Blocking		Simultaneous Blocking		Relative Validity	
Phase I A+: 36	Phase II AW+: 36	Phase I A+:12 B-:12 -:12	Phase II AW+:9 BX+:9 YZ+:9 -:9	A+:18 AW+:18	A+:18 AW+:18

B. Simulation of what would Happen in the First Third of the Session (12 Trials) after the First Occurrence of the Target Cue, W

Blocking		Simultaneous Blocking		Relative Validity	
Phase I A+: 36	Phase II AW+: 12	Phase I A+:12 B-:12 -:12	Phase II AW+:3 BX+:3 YZ+:3 -:3	A+:6 AW+:6	
Total: 48		Total: 48		Total: 12	

C. Estimation of the Relevance Relativization from PCPM for the Alternative Cue A, which would also be Included in the Computations for Stimulus W

Blocking	Simultaneous Blocking	Relative Validity
$\left\{ \begin{array}{l} p(A \cap W) = 12 / 48 = 0.25 \\ p(A \cap \bar{W}) = 36 / 48 = 0.75 \end{array} \right\}$	$\left\{ \begin{array}{l} p(A \cap W) = 3 / 48 = 0.06 \\ p(A \cap \bar{W}) = 12 / 48 = 0.25 \end{array} \right\}$	$\left\{ \begin{array}{l} p(A \cap W) = 6 / 12 = 0.50 \\ p(A \cap \bar{W}) = 6 / 12 = 0.50 \end{array} \right\}$
$ p(A \cap W) - p(A \cap \bar{W}) = 0.50$	$ p(A \cap W) - p(A \cap \bar{W}) = 0.19$	$ p(A \cap W) - p(A \cap \bar{W}) = 0.00$

D. Predictions for the Target Cue, W

Blocking	Simultaneous Blocking	Relative Validity
Predictive Context: No PMPC: No >Pr/Ca Judg to W: Intermediate Causal Context: Yes PMPC: No >Pr Judg to W: Intermediate >Ca Judg to W: Null (only with tests apt to foster competition & Pr. Model)	Predictive Context: No PMPC: Tends to Yes >Pr/Ca Judg to W: Intermediate Causal Context: Yes PMPC: Tends to Yes >Pr Judg to W: Intermediate >Ca Judg to W on Pr Model: Null (almost always) >Ca Judg to W on Dg Model: Intermediate	Predictive Context: No PMPC: Yes >Pr/Ca Judg to W: Null (only with tests apt to foster competition & Pr. Model) Causal Context: Yes PMPC: Yes >Pr Judg to W: Intermediate >Ca Judg to W on Pr Model: Null >Ca Judg to W on Dg Model: Null (only with tests apt to foster competition between Effects)

Note: Numbers within charts are the frequencies of each event. In section D, type of context (Predictive or Causal), result of PCPM, judgment type (Pr/Ca Judg, Predictive or Causal judgment), and causal model (Predictive, Pr, or Diagnostic, Dg).

first panel displays the designs of the three tasks. The second one shows the estimation of what would happen in the first third of the sessions after the appearance of the potentially overshadowed cue (W), that is, we simulate what would happen after 12 trials following the presence of this stimulus. Finally we detail the estimations of the two probabilities involved in the estimation of the PMPC together with the computation of PMPC. The last panel displays the predictions for the target cue derived from our model. As can be seen in panel C, blocking and relative validity are two extreme cases of PMPC (.50 and .00 respectively) and simultaneous blocking (.19) is an intermediate case, nearer to relative validity. We predict that the Causal/Predictive judgment about the target cue (W) will be of an intermediate magnitude in the Blocking design (remember that the raw validity of this stimulus is .50, as it only appeared in the middle of the trials sequence), and around 0 in Relative Validity (when validity will be 0, as in the example of Fig. 2).

However, the type of scenario (Predictive or Causal), the demands imposed by the different types of judgments (Predictive or Causal), together with the type of Causal Model (Predictive or Diagnostic) in causal situations can modulate these predictions, as can be derived from the two other principles: RR-by-Causality and RR-by-Mental-Model (see Section D in Table 1). The most representative case is a predictive scenario in which relativization would not take place when coding the information but would intervene when translating what was learned into performance. This is where the PMPC would be used, so that the final response would in fact be affected by the competition principle, making the predictive judgment null although it was coded with an intermediate value. This happens when the test question fosters competition and the mental model is diagnostic.

AUTHOR NOTE

This research was financially supported by Junta de Andalucía, Spain, Research Grant HUM434 to A. R. Ortega, Grant HUM642 to J.M. Rosas, Grant PB06-02375 to A. Catena; and Ministerio de Ciencia y Tecnología, Spain, Grant BSO2002-03398 to J.M. Rosas. Correspondence concerning to this article should be addressed to Manuel Miguel Ramos Álvarez, Departamento de Psicología, Universidad de Jaén, 23071 Jaén, Spain. E-mail: mramos@ujaen.es

REFERENCES

- [1] Cheng PW, Holyoak KJ. Complex adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In: Meyer JA, Roitblat H, Eds. *Comparative approaches to cognition*. Cambridge, MA: MIT Press 1995; pp. 271-302.
- [2] Kareev Y. Positive bias in the perception of covariation. *Psychol Rev* 1995; 102: 490-502.
- [3] Kareev Y. Seven (indeed, plus or minus two) and the detection of correlations. *Psychol Rev* 2000; 107(2): 397-402.
- [4] Cheng PW. Separating causal laws from casual facts: pressing the limits of statistical relevance. In: Medin DL, Ed. *The Psychology of Learning and Motivation*. San Diego: Academic Press 1993; vol. 30: pp. 215-264.
- [5] Cheng PW, Novick LR. Covariation in natural causal induction. *Psychol Rev* 1992; 99(2): 365-82.
- [6] Shanks DR. Human Instrumental Learning: a critical review of data and theory. *Br J Psychol* 1993; 84: 319-54.
- [7] Shanks DR. Human associative learning. In: Mackintosh NJ, Ed. *Handbook of perception and cognition*. Animal learning and cognition. San Diego, CA: Academic Press 1994; vol. 9: pp. 335-74.
- [8] Young ME. On the origin of personal causal theories. *Psychon Bull Rev* 1995; 2: 83-104.
- [9] Salmon WC. *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press 1984.
- [10] Suppes PA. *Probabilistic theory of causality*. Amsterdam: North-Holland 1970.
- [11] De Houwer J, Beckers T. A review of recent developments in research and theories on human contingency learning. *Q J Exp Psychol B* 2002; 55: 289-310.
- [12] Spellman BA. Acting as intuitive scientists: contingency judgments are made while controlling for alternative potential causes. *Psychol Sci* 1996; 7: 337-42.
- [13] Buehner M, Cheng P. Causal Learning. In: Holyoak KJ, Morrison RG, Eds. *The Cambridge handbook of thinking and reasoning*. NY: Cambridge University Press 2005; pp. 143-68.
- [14] Allan LG, Tangen JM. Judging relationships between events: How do we do it? *Can J Exp Psychol* 2005; 59: 22-7.
- [15] Mitchell CJ, De Houwer J, Lovibond PF. The propositional nature of human associative learning. *Behav Brain Sci* 2009; 32: 183-246.
- [16] Pineño O, Denniston JC, Beckers T, Matute H, Miller RR. Contrasting predictive and causal values of predictors and of causes. *Learn Behav* 2005; 33(2): 184-96.
- [17] Cheng PW. From covariation to causation: a causal power theory. *Psychol Rev* 1997; 104: 367-405.
- [18] Spellman BA. Conditionalizing causality. In: Shanks DR, Holyoak KJ, Medin DL, Eds. *The Psychology of Learning and Motivation*. Causal learning. San Diego: Academic Press 1996; vol. 34: pp.167-206.
- [19] Shanks DR, Holyoak KJ, Medin DL, Eds. *The Psychology of Learning and Motivation*. Causal learning. San Diego: Academic Press 1996; vol. 34.
- [20] Wickens TD. *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum 1989.
- [21] Larkin MJW, Aitken MRF, Dickinson A. Retrospective reevaluation of causal judgments under positive and negative contingencies. *J Exp Psychol Learn Mem Cogn* 1998; 24(6): 1331-52.
- [22] Matute H, Arcediano F, Miller RR. Test question modulates cue competition between causes and between effects. *J Exp Psychol Learn Mem Cogn* 1996; 22(1): 182-96.
- [23] Chapman GB, Robbins SJ. Cue interaction in human contingency judgment. *Mem Cognit* 1990; 18(5): 537-45.
- [24] Price PC, Yates, JF. Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Mem Cogn* 1993; 21(5): 561-72.
- [25] Waldmann MR. Competition among causes but not effects in predictive and diagnostic learning. *J Exp Psychol Learn Mem Cogn* 2000; 26: 53-76.
- [26] Shanks DR, Lopez FJ. Causal order does not affect cue selection in human associative learning. *Mem Cogn* 1996; 24: 511-22.
- [27] Waldmann MR. Knowledge-based causal induction. In: Shanks DR, Holyoak KJ, Medin DL, Eds. *The Psychology of Learning and Motivation*. Causal learning. San Diego: Academic Press 1996; vol. 34: pp. 47-88.
- [28] White PA. Beliefs about interactions between factors in the natural environment: A causal network study. *Appl Cogn Psychol* 2008; 22: 559-72.
- [29] Morris MW, Larrick RP. When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychol Rev* 1995; 102(2): 331-55.
- [30] Novick LR, Cheng PW. Assessing interactive causal influence. *Psychol Rev* 2004; 111: 455-85.
- [31] Baker AG, Singh M, Bindra D. Some effects of contextual conditioning and US predictability on Pavlovian conditioning. In: Balsam PD, Tomie A, Eds. *Context and learning*. Hillsdale, NJ: Erlbaum 1985; pp. 73-103.
- [32] Kirk RE. *Experimental designs: procedures for the behavioral sciences*, 2nd ed. Belmont, CA: Brooks and Coleman 1982.
- [33] Bussemeyer JR, Myung IJ, McDaniel MA. Cue competition effects: Theoretical implications for adaptive network learning models. *Psychol Sci* 1993; 4(3): 196-202.
- [34] Estes WK. The cognitive side of probability learning. *Psychol Rev* 1976; 83: 37-64.
- [35] Dudycha LW, Naylor JC. Characteristics of the human inference process in complex choice behaviour situations. *Organ Behav Hum Perform* 1966; 1: 110-28.

- [36] Catena A, Maldonado A, Candido A. The effect of the frequency of judgment and the type of trials on covariation learning. *J Exp Psychol Hum Percept Perform* 1998; 24: 481-95.
- [37] Hogarth RM, Einhorn HJ. Order effects in belief updating: the belief-adjustment model. *Cogn Psychol* 1992; 24: 1-55.
- [38] Cooksey RW. Judgment analysis. Theory, methods, and applications. San Diego, CA: Academic Press 1996.
- [39] Busemeyer JR. Intuitive statistical estimation. In: Anderson NH, Ed. Contributions to information integration theory. Cognition. Hillsdale, NJ: Erlbaum 1991; vol. 1: pp. 187-215.
- [40] McKenzie CRM. The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cogn Psychol* 1994; 26: 209-39.
- [41] Van Hamme LJ, Wasserman EA. Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learn Motiv* 1994; 25: 127-51.
- [42] Catena A, Maldonado A, Perales JC, Candido A. Interaction between previous beliefs and cue predictive value in covariation-based causal induction. *Acta Psychol (Amst)* 2008; 128(2): 339-49.
- [43] Allport A. Visual attention. In: Posner M, Ed. Foundations of cognitive science. Cambridge, MA: MIT Press 1989; pp. 631-82.
- [44] Kaufman MA, Bolles RC. A nonassociative aspect of overshadowing. *Bull Psychon Soc* 1981; 18: 318-20.
- [45] Miller RR, Barnet RC, Grahame NJ. Responding to a conditioned stimulus depends on the current associative status of other cues present during training of that specific stimulus. *J Exp Psychol Anim Behav Process* 1992; 18: 251-64.
- [46] Bindra D, Clarke KA, Shultz TR. Understanding predictive relations of necessity and sufficiency in formally equivalent "causal" and "logical" problems. *J Exp Psychol Gen* 1980; 109: 422-43.
- [47] Matute H, Vegas S, De Marez P. Flexible use of recent information in causal and predictive judgments. *J Exp Psychol Learn Mem Cogn* 2002; 28: 714-25.
- [48] Wagner AR. Stimulus selection and a "modified continuity theory". In: Bower GH, Spence JT, Eds. *The Psychology of Learning and Motivation*, San Diego, CA: Academic Press 1969; Vol. 3. pp. 1-43.
- [49] Kamin LJ. "Attention-Like" processes in classical conditioning. In: Jones MR, Ed. *Miami symposium on the prediction of behavior: aversive stimulation*. Miami: University of Miami Press 1968; pp. 9-33.
- [50] Kamin LJ. Predictability, surprise, attention and conditioning. In: Campbell B, Church R, Eds. *Punishment and aversive behavior*. New York: Appleton-Century-Crofts 1969; pp. 279-96.
- [51] Ramos-Alvarez MM, Catena A. Lo que se aprende sobre la predicción de una consecuencia no tiene por qué influir sobre otros predictores potenciales. In: Ramos MM, Berrios P, Eds. *Investigación en Psicología*. Jaén: Servicio de Publicaciones de la Universidad de Jaén 2007; vol. 1: pp. 105-40.
- [52] Allan LG. A note on measurement of contingency between two binary variables in judgment tasks. *Bull Psychon Soc* 1980; 15: 147-49.
- [53] Domjan M, Burkhard B. *The principles of learning and behavior*. CA: Brooks/Cole Pub Co 1992.
- [54] Tangen JM, Allan LG. Cue interaction and judgments of causality: Contributions of causal and associative processes. *Mem Cogn* 2004; 32(1): 107-24.
- [55] Shanks DR. On similarities between causal judgments in experienced and described situations. *Psychol Sci* 1991; 2(5): 341-50.
- [56] Judd ChM, McClelland GH. *Data analysis. A model-comparison approach*. San Diego: Harcourt Brace Jovanich 1989.
- [57] Allan LG, Hannah SD, Crump MJC, Siegel S. The psychophysics of contingency assessment. *J Exp Psychol Gen* 2008; 137: 226-43.
- [58] Wasserman EA, Elek SM, Chatlosh DL, Baker AG. Rating causal relations: role of probability in judgments of response-outcome contingency. *J Exp Psychol Learn Mem Cogn* 1993; 19(1): 174-88.
- [59] Chapman GB. Trial order affects cue interaction in contingency judgment. *J Exp Psychol Learn Mem Cogn* 1991; 17(5): 837-54.
- [60] Hannah SD, Crump MJC, Allan LG, Siegel S. Cue-interaction effects in contingency judgments using the streamed-trial procedure. *Can J Exp Psychol* 2009; 63: 103-12.
- [61] Baker AG, Mercier P, Vallée-Tourangeau F, Frank R, Pan M. Selective associations and causality judgments: presence of a strong causal factor may reduce judgments of a weaker one. *J Exp Psychol Learn Mem Cogn* 1993; 19: 414-32.
- [62] Baetu I, Baker AG, Darreseau C, Murphy RA. A comparative approach to cue competition with one and two strong predictors. *Learn Behav* 2005; 33(2): 160-71.
- [63] Van Hamme LJ, Kao SF, Wasserman EA. Judging interevent relations: From cause to effect and from effect to cause. *Mem Cogn* 1993; 21(6): 802-8.
- [64] Williams DA, Sagness KE, McPhee JE. Configural and elemental strategies in predictive learning. *J Exp Psychol Learn Mem Cogn* 1994; 21: 129-42.

Received: October 07, 2009

Revised: January 14, 2010

Accepted: January 18, 2010

© Ramos-Álvarez and Catena; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License

(http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.