THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (5), 877–889

Psychology Press
Taylor & Francis Group

Generality of the summation effect in human causal learning

Fabian A. Soto, Edgar H. Vogel, and Ramón D. Castillo
Universidad de Talca, Talca, Chile

Allan R. Wagner
Yale University, New Haven, CT, USA

Considerable research has examined the contrasting predictions of the elemental and configural association theories proposed by Rescorla and Wagner (1972) and Pearce (1987), respectively. One simple method to distinguish between these approaches is the summation test, in which the associative strength attributed to a novel compound of two separately trained cues is examined. Under common assumptions, the configural view predicts that the strength of the compound will approximate to the average strength of its components, whereas the elemental approach predicts that the strength of the compound will be greater than the strength of either component. Different studies have produced mixed outcomes. In studies of human causal learning, Collins and Shanks (2006) suggested that the observation of summation is encouraged by training, in which different stimuli are associated with different submaximal outcomes, and by testing, in which the alternative outcomes can be scaled. The reported experiments further pursued this reasoning. In Experiment 1, summation was more substantial when the participants were trained with outcomes identified as submaximal than when trained with simple categorical (presence/absence) outcomes. Experiments 2 and 3 demonstrated that summation can also be obtained with categorical outcomes during training, if the participants are encouraged by instruction or the character of training to rate the separately trained components with submaximal ratings. The results are interpreted in terms of apparent performance constraints in evaluations of the contrasting theoretical predictions concerning summation.

Keywords: Summation; Configural; Elemental; Causal learning; Rescorla–Wagner model.

One of the most controversial issues in modern theories of associative learning is whether the associated stimuli are processed in elemental or configural fashion. Elemental theories, such as that of Rescorla and Wagner (1972), propose that associative strength is acquired and expressed to each of the separable components of the stimulus complex, whereas configural theories, such as that of Pearce (1987), propose that associative strength is acquired and expressed to entire stimulus configurations.

Certain phenomena are more agreeable with the elemental approach, like the fact that the response tendencies of cues trained separately "summate" when the cues are presented in compound. The Rescorla and Wagner (1972) model assumes that

Correspondence should be addressed to Edgar Vogel, Universidad de Talca, Escuela de Psicología, Casilla N° 747, Talca, Chile. E-mail: evogel@utalca.cl

This work was supported by Fondecyt Grant 1060838 to E. H. Vogel.

http://www.psypress.com/qjep
DOI:10.1080/17470210802373688

the associative strength of a compound is equal to the algebraic sum of the associative strength of its components. It thus predicts that the associative strength of the novel compound in a summation test will be equivalent to the sum of the associative strengths acquired by each element, and that responding to the compound should be greater than that to each of the separate stimuli.

Conversely, a basic configural approach, like Pearce's (1987) theory, without qualification, predicts no summation. The assumption is that compound stimuli are processed as unique exemplars and form associations independently of those that might be formed by their component elements. The component elements do, however, determine the generalization that occurs between similar configurations. According to the generalization rule adopted by Pearce (1987), the proportion of associative strength generalized from configuration j to configuration i is computed as the proportion of stimulus elements in the first configuration that are common to both, multiplied by the proportion of stimulus elements in the second configuration that are common to both ($N_C/N_i \times N_C/N_j$, where N_C stands for the number of elements common to both configurations, and N_i and N_j stand for the number of elements in configurations i and j, respectively). In a summation test, each trained element, A and B, provides half of their respective acquired associative strength to the novel compound AB, so that responding to the compound is predicted to approximate to the mean of the associative strength of the elements A and B.

In the Pavlovian conditioning literature, studies testing summation have produced a variety of results. A substantial literature involving rats and rabbits, trained in a number of different circumstances, using stimuli from separate modalities, has reported summation (Kehoe, Horne, Horne, & Macrae, 1994; Myers, Vogel, Shin, & Wagner, 2001; Rescorla, 1997; Whitlow & Wagner, 1972; but see Pearce, George, & Aydin, 2002). In contrast, studies of summation using pigeon autoshaping and stimuli within the visual modality have consistently reported the failure of summation (Aydin & Pearce, 1994, 1995, 1997; Rescorla & Coldwell, 1995). The different results have led to different proposals for how they could be reconciled with elemental and configural theories. For example, Wagner (2003) has suggested that the critical difference is in the stimuli involved, with more separable stimuli producing summation and more integral stimuli not. Alternatively, Pearce (2002) has suggested that the critical difference may be in the influence of contextual stimuli that need be considered to be in compound with the training and testing cues, with summation failing in the absence of such cues, but occurring with salient contextual cues. We return to these interpretations in the Discussion section.

A small number of studies have examined summation in human causal learning. In one of these studies, van Osselaer, Janiszewski, and Cunha (2004) trained college students to predict different levels of quality of bottled wines that differed in a number of features. Following training on an A+, B+, CD+ task, responding to a novel AB compound was greater than that to CD, an outcome that is consistent with elemental theory. Unfortunately, these authors did not report the magnitude of responding to A and B individually.

More recently, Collins and Shanks (2006) provided more convincing evidence of elemental processing, demonstrating summation to an AB compound relative to its A and B elements presented individually. Participants were asked to learn the causal relationship between several radiation types and different levels of mutation in DNA samples. During the first phase of two experiments, two different radiation types were followed by the same submaximal level of DNA mutation (in a scale that ranged from 0 to 80, Cues A and B produced a mutation level of 20, and Cues E and F produced a mutation level of 60), and the participants had to estimate the level of DNA mutation that each stimulus would produce in a scale with endpoints labelled "low" and "high". When presented with the same individual cues and their compounds in a further experimental phase, the mean estimation of DNA mutation assigned to the compounds was reliably higher than that for their respective component cues.

Interestingly, these two studies that have found summation in human casual learning (Collins &

Shanks, 2006; van Osselaer et al., 2004) used a task in which the outcome during training was quantitative and submaximal. Collins and Shanks (2006) argued that training with submaximal magnitude outcomes may encourage participants to quantitatively differentiate between the causal strength of compounds and its components, and that this is less likely to happen after training with maximal categorical outcomes. Furthermore, they argued that causal judgements should be assessed using a continuum of magnitudes, allowing participants to reflect the differences in their estimation of causal strength for different events.

The hypothesis that the nature of the outcomes identified in training (e.g., magnitude vs. categorical; maximal vs. submaximal) is a critical variable in producing elemental-like results has not been directly evaluated. The goal of the present experiments was to comment on this possibility. In Experiment 1, separate groups of college students were trained in a food-allergy prediction task with categorical outcomes that were either of unspecified magnitude (allergy vs. no allergy) or of submaximal magnitude (0 points vs. 10 points of allergy, out of a total of 20 points). When training was with submaximal outcomes the test response to the training stimuli was lessened, and the response to the stimuli in compound showed clearer summation. Experiments 2 and 3 were designed to further evaluate the summation effect after training with categorical outcomes, using procedures that otherwise facilitated submaximal ratings of the trained elements during test. In Experiment 2, participants were explicitly asked during test to use submaximal values to rate the causal strength of the stimuli paired with the outcome in training. In Experiment 3, the submaximal ratings of the to-be-compounded stimuli, A and B, were achieved by their training in compound with other cues (i.e., AC+ and BD+). Again, summation was evident under these conditions.

EXPERIMENT 1

In the task employed by Collins and Shanks (2006), different stimuli were paired in training with different degrees of DNA mutation, all of which were submaximal. In testing, participants were asked to judge the degree of mutation associated with the different stimuli and their combinations. It may be critical to the observance of summation that participants had experience with different degrees of outcome and had opportunity to judge the outcomes on the scale with which they were trained. But it is also possible that identification of the outcomes associated with the training elements as submaximal, or the testing of different degrees of outcome, is sufficient to encourage the observance of summation in their novel combination. In Experiment 1 summation was tested under three conditions in separate groups, in an allergy-predicting context. In group qualitative/likelihood, training was with outcomes of unspecified magnitude (allergic reaction vs. no allergic reaction), and testing involved judgements of the likelihood of allergic reaction to the test stimuli, on a scale of 0 to 20. In group qualitative/intensity, training was similarly with outcomes of unspecified magnitude, but testing involved judgements of the intensity of allergic reaction on a scale of 0 to 20. In group quantitative/intensity training was with outcomes specified as 0 or 10 points of allergy on a scale of 0 to 20, and testing involved judgements of the intensity of reactions on the same scale.

Training (see Table 1) involved two single cues that were independently said to be followed by an allergic reaction (A+, B+), two single cues not

Table 1. *Design of Experiment 1*

Group	*Training*	*Test*
Qualitative/likelihood	Stimuli	A
	A+	B
	B+	EF
Qualitative/intensity	EF+	AB
	Fillers	
Quantitative/intensity	C−	C
	D−	D
	GH−	GH
		CD

Note: Letters A–H represent different foods that could be followed (+) or not followed (−) by an allergic reaction in a hypothetical patient.

followed by an allergic reaction (C− and D−), one compound cue followed by an allergic reaction (EF+), and one compound cue not followed by an allergic reaction (GH−). The negative cues were fillers that forced the participants to discriminate which foods were followed by an allergic reaction and which were not. In test, they were presented again with all trained cues, A, B, C, D, EF, and GH, as well as the novel AB and CD compounds.

Method

Participants

A total of 48 undergraduate students at the University of Talca participated in the experiment for course credit. They were tested individually and had no previous experience in similar research. Participants were randomly assigned to one of three groups ($n = 16$).

Materials

In the training phase, stimuli were presented, and data were collected with a HP Compaq personal computer connected to a 14-inch colour screen and programmed with E-prime software (version 1.1; Psychology Software Tools, Inc., Pittsburgh, PA). In the testing phase, causal judgements were assessed by asking the participants to fill out a rating sheet by pencil.

Procedure

The experimenter told the participants that all necessary instructions would be presented on the "instruction screens" included in the computer program. They then were left to complete the experiment in a private room and were asked to inform the experimenter when they finished the experiment.

At the beginning of the training phase the following instructions appeared on the screen (in Spanish):

In this experiment we will ask you to imagine that you are an allergist (someone who tries to discover the causes of allergic reactions in people). Imagine that just now a new patient arrives, Mr. X, who suffers from allergic reactions after eating some foods, but not others. In an attempt to discover which foods cause allergic reactions in Mr. X, you ask him to eat several foods for a meal on each day, and observe if he had an allergic reaction or not.

The computer will show you the foods that Mr. X eats in each meal. Next you will be asked to predict whether or not Mr. X will have an allergic reaction. Enter your prediction by pressing the letter "a" in the keyboard to indicate that Mr. X will have an allergic reaction after eating that meal, or pressing "n" to indicate that Mr. X will not have an allergic reaction after eating the meal. A message will appear in the right side of the screen indicating whether or not an allergic reaction actually occurred. You will have to guess at first, but with the aid of the feedback your predictions should soon start to become more accurate.

Your reaction times are not important in this experiment. You may take as long as you like on each trial.

PLEASE PRESS THE SPACE BAR TO CONTINUE

A series of 84 trials was presented to the participant. At the beginning of each trial, a stimulus or pair of stimuli was shown on the left-centre portion of the screen, followed by the phrase, "Press 'a' to indicate an allergic reaction, and 'n' to indicate no allergic reaction" at the bottom of the screen. After the participant entered a response, feedback was provided on the right side of the screen for 2 seconds. The feedback consisted of the phrase "Allergic reaction" in red, size 30 Arial font for the A+, B+ and EF+ trials, or the phrase "No allergic reaction" in black, size 14 Arial font for the C−, D−, GH− trials. This was the only feedback information given to participants in groups qualitative/likelihood and qualitative/intensity. For participants in the group quantitative/intensity, additional quantitative feedback information was provided with the sentence "0 points of allergy over a total of 20" or "10 points of allergy over a total of 20", which appeared underneath the phrases "no allergic reaction" and "allergic reaction", respectively.

A total of 14 presentations of each trial type occurred in random order for each participant. The assignment of specific foods to the conditions A−H was partially counterbalanced across participants by means of their different allocation in one of eight subgroups, each with a different assignment of foods as A−H. Specifically, in Subgroup 1 the cues A to H were chocolate, milk, cheese, garlic, peanuts, lobster, coffee, and sardines, respectively. To create Subgroup 2, each food was moved forward one position in the list

(i.e., chocolate was assigned to B, milk to condition C, etc., and sardines was assigned to A). Subgroups 3 to 8 were obtained by iterating this procedure. The position (right vs. left) of the stimuli forming a compound was counterbalanced across the experiment. That is, in half of the trials the stimuli were presented in one position (e.g., EF), and in the other the relative position was reversed (e.g., FE).

Upon completion of the 84 training trials, the participants were presented with the following message at the top of the screen: "You have finished your examination of Mr. X and now must fill out a diagnosis report. For this, please complete the questionnaire that is in the drawer next to the computer. You will find the instructions at the beginning of the questionnaire."

The printed questionnaire was two pages long. On the first page, the following instructions were presented to participants in group qualitative/likelihood.

Now we will ask you to predict the consequences to Mr. X of eating various meals. Please, base your predictions on the information you learned about which foods provoke allergic reactions to Mr. X and which foods do not.

These meals can contain one food or two. You have to judge how likely Mr. X is to have an allergic reaction following the meal (i.e., if the meal contains two foods, you should rate how likely Mr. X is to have an allergic reaction after eating both foods together).

To rate the effect of each meal, use a scale of 0 to 20 points, where 0 points mean that eating the meal is very unlikely to cause an allergic reaction in Mr. X and 20 points means that eating the meal is very likely to cause an allergic reaction in Mr. X. Use intermediate values to indicate different probabilities of allergy in between the two extremes.

To enter your rating, mark your choice. If you make a mistake or want to change your rating, you can erase the previously selected alternative and make a new choice.

When you finish, give this booklet to the experimenter. Thank you for your participation in this experiment.

Participants in groups qualitative/intensity and quantitative/intensity received similar instructions to those described above, but every reference to "likelihood of allergic reaction" was replaced with references to "intensity of allergic reaction". The test instructions for these two groups were as follows:

Now we will ask you to predict the consequences to Mr. X of eating various meals. Please, base your predictions on the information you learned about which foods provoke allergic reactions to Mr. X and which foods do not.

These meals can contain one food or two. You have to judge the level of allergic reaction Mr. X will experience following the meal (i.e., if the meal contains two foods, you should rate the intensity of allergic reaction experienced by Mr. X after eating both foods together).

To rate the effect of each meal, use a scale of 0 to 20 points, where 0 points mean that eating the meal will cause no allergic reaction to Mr. X and 20 points means that eating the meal will cause to Mr. X an allergic reaction of maximal intensity. Use intermediate values to indicate different degrees of allergy in between the two extremes.

To enter your rating, mark your choice. If you make a mistake or want to change your rating, you can erase the previously selected alternative and make a new choice.

When you finish, give this booklet to the experimenter. Thank you for your participation in this experiment.

The second page of the questionnaire was identical for the three groups and included a list of the eight test stimuli outlined in Table 1. Next to each food there were 21 circles with numbers inside ranging from 0 to 20. Test stimuli were ordered in the questionnaire in a different, semirandom order for each participant.

Results and discussion

One participant in each group failed to rate all eight cues of the booklet, and their data were discarded from the analyses.

Figure 1 presents the mean causal ratings for the individually trained cues (A and B), for the trained compound (EF), and for the novel compound formed by the individually trained cues (AB), in the three groups of Experiment 1. The data depicted suggest that summation, in the form of more responding to the novel AB compound than to the previously trained cues A, B, and EF, was observed to some degree in all three groups. They further suggest that the level of summation in AB was substantial in group quantitative/intensity, modest in group qualitative/intensity, and weak, if at all, in group qualitative/likelihood. Finally, all negative cues were judged with very low causal strengths, which indicate that the participants learned the discrimination.

A 3 (group) × 8 (cues: A, B, EF, AB, C, D, GH, CD) mixed design analysis of variance (ANOVA)

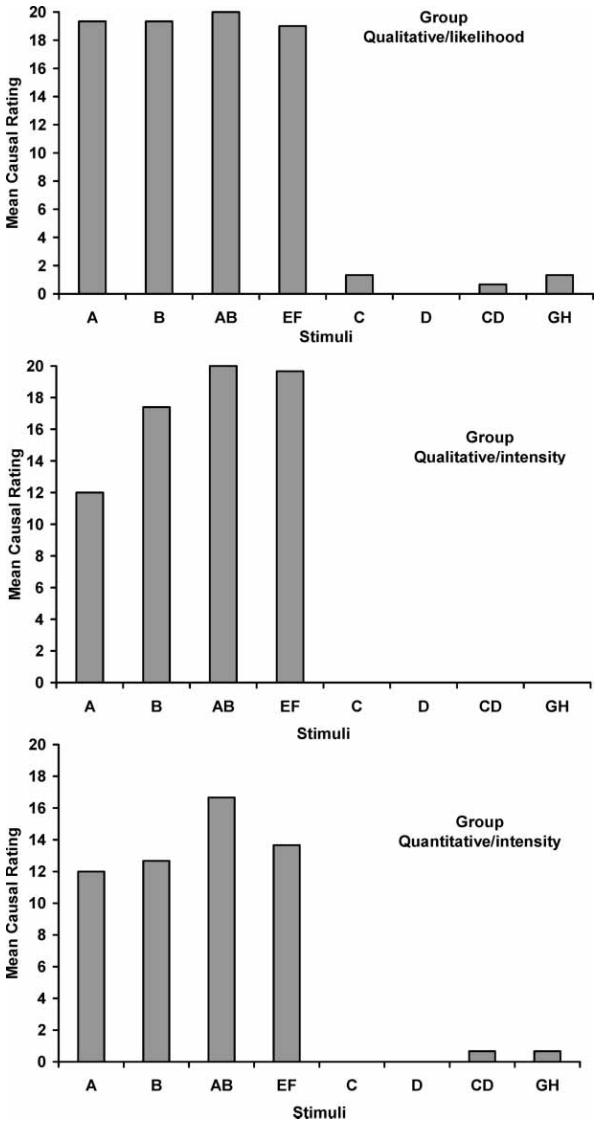

Figure 1. *Mean causal ratings to the relevant stimuli in the summation test of Experiment 1. The results are showed separately for groups qualitative/likelihood, qualitative/intensity, and quantitative/intensity.*

indicate reliable main effects of group, $F(2, 42) = 16.130$, $p < .001$, partial $\eta^2 = .434$, and cues, $F(7, 294) = 510.101$, $p < .001$, partial $\eta^2 = .924$, and a reliable interaction between group and cues, $F(14, 294) = 6.369$, $p < .001$, partial $\eta^2 = .233$.

The main effect of group was due to the lower ratings in group quantitative/intensity, in which the participants were trained with an allergy value of 10/20 for the cues A, B, and EF, in comparison to the other two groups, which were trained with unspecified allergy value. This was confirmed by least significant difference (LSD) post hoc comparisons, which indicate that the ratings of group quantitative/intensity were significantly lower than those of the group qualitative/likelihood, $t(42) = -5.481$, $p < .001$, partial $\eta^2 = .518$, and of the group qualitative/intensity, $t(42) = -4.030$, $p < .001$, partial $\eta^2 = .367$, which, in turn, did not differ from each other, $t(42) = 1.452$, $p = .154$, partial $\eta^2 = .070$.

The Group × Cues interaction was examined by ANOVAs testing the simple main effect of cues within each group, followed by pairwise LSD post hoc tests. This analysis indicated no reliable evidence of summation in the group qualitative/likelihood since the rating for AB did not differ significantly from those for A, $t(42) = 0.651$, $p = .518$, partial $\eta^2 = .010$, B, $t(42) = 0.658$, $p = .514$, partial $\eta^2 = .010$, and EF, $t(42) = 1.220$, $p = .229$, partial $\eta^2 = .034$. Some evidence of summation was seen in the group qualitative/intensity, where the ratings assigned to AB were significantly higher than those to A, $t(42) = 2.540$, $p = .015$, partial $\eta^2 = .133$, and those to B, $t(42) = 2.566$, $p = .014$, partial $\eta^2 = .136$, but did not reliably differ from the ratings assigned to EF, $t(42) = 0.407$, $p = .686$, partial $\eta^2 = .004$. Finally, reliable summation was demonstrated in group quantitative/intensity, where the ratings to cue AB were significantly higher than those to each of A, $t(42) = 4.558$, $p < .001$, partial $\eta^2 = .331$, B, $t(42) = 3.947$, $p < .001$, partial $\eta^2 = .271$, and EF, $t(42) = 3.660$, $p = .001$, partial $\eta^2 = .242$, which, in turn, did not differ reliably from each other (ps $> .091$, partial η^2s $< .67$).

The apparent summation effect observed in group quantitative/ intensity cannot be accounted for by assuming that novel compounds were judged more allergenic than familiar compounds, since the response to the novel compound CD did not differ significantly from those to C, $t(42) = 1.225$, $p = .227$, partial $\eta^2 = .034$, D, $t(42) = 1.225$, $p = .227$, partial $\eta^2 = .034$, and GH, $t(42) = 0.000$, in this group.

Finally, a further, more conservative, test of summation was conducted, in which responding to AB was compared across subjects with the responding to whichever was the more responded to of A or B alone for each subject. This analysis showed again evidence of summation in group quantitative/intensity since the response to AB was reliably greater than the greater of A or B of each subject, $t(14) = 3.055$, $p = .009$, partial $\eta^2 = .40$.

These results provide direct evidence that although summation can be observed in a human causal learning preparation, there are training conditions that do not favour it. The failure to observe reliable summation in groups qualitative/likelihood and qualitative/intensity, as contrasted to that observed in group quantitative/intensity, supports the hypothesis proposed by Collins and Shanks (2006) that detecting summation may require training and testing procedures that allow participants to acknowledge that the outcome associated with the training elements alone is submaximal. However, it should be noted that doing this need not require that the participants have experience with different submaximal outcomes, or with different quantitative values for some compounds versus elements, as in the Collins and Shanks experiments. It was sufficient in Experiment 1 that participants had the outcome in training identified as a submaximal 10 out of the possible 20.

It is notable that the magnitude of the trend toward summation in the several groups of Experiment 1 was a function of the level to which causal ratings assigned to the individual cues A and B was below the ceiling of the measurement scale. It is possible that the observation of reliable summation in group quantitative/intensity was due simply to the procedures in that group affording an appropriately low level of response to A and B in testing. If so, it is possible that a summation effect could be obtained by pairing the individual stimuli with a binary outcome during training and testing with a likelihood scale, as in the case of group qualitative/likelihood, if the experimental procedure were otherwise modified to facilitate the rating of the individual stimuli A and B with submaximal values during the summation test. This was accomplished in different ways in Experiments 2 and 3.

EXPERIMENT 2

The design of this experiment was essentially the same as that for group qualitative/likelihood of Experiment 1, which showed very high levels of response to all of the positive stimuli and no evidence of summation. The critical difference was an additional set of instructions provided during test, in which the participants were asked to rate with submaximal values the likelihood of allergic reaction to the various foods that caused allergic reaction during training. Care was taken to indicate that novel meals that might be present in testing could be rated at a lower level, the same level, or a higher level.

Method

Participants

A total of 24 undergraduate students at the University of Talca participated in the experiment for course credit. They were tested individually and had no previous experience in similar research.

Materials

All the features of the materials and procedure were the same as those for group qualitative/likelihood of Experiment 1, except for the instructions in the printed questionnaire that was used in the testing phase. The printed questionnaire was two pages long. On the first page, the following instructions were presented to the participants:

Now we will ask you to predict the consequences to Mr. X of eating various meals. Please, base your predictions on the information you learned about which foods provoke allergic reactions to Mr. X and which foods do not.

These meals can contain one food or two. You have to judge how likely Mr. X is to have an allergic reaction following the meal (i.e., if the meal contains two foods, you should rate how likely Mr. X is to have an allergic reaction after eating both foods together).

To rate the effect of each meal, use a scale of 0 to 20 points. During the previous examination, you have already observed

the effect that some of these meals have on Mr. X. Based on what you remember, rate as 10 points those meals that you believe probably will cause an allergic reaction in Mr. X, and with 0 points those meals that probably will not cause an allergic reaction in Mr. X.

Furthermore, we will ask you to rate some meals that did not appear in the previous examination. These meals are not completely new, but could be combinations of two foods that Mr. X ate separately in the examination. It is possible that you believe that these new meals are more or less likely to produce an allergic reaction in Mr. X than those that caused an allergic reaction during the examination. If you believe that the new meal is **more likely** to produce an allergy than those previously seen, rate it with values **above 10 (between 11 and 20)**. If you believe that the new meal is **less likely** to produce an allergic reaction in Mr. X than those previously seen, rate it with values below **10 (between 0 and 9).** Finally, rate the new meal with 10 if you think that it is equally likely to cause allergic reaction in Mr. X than the foods that produced it during the examination.

To enter your rating, mark your choice. If you make a mistake or want to change your rating, you can erase the previously selected alternative and make a new choice.

When you finish, give this booklet to the experimenter. Thank you for your participation in this experiment.

The second page of the questionnaire included a list of the 8 test stimuli outlined in Table 1. Next to each food there were 21 circles with numbers inside ranging from 0 to 20. Test stimuli were presented in a different, semirandom order for each participant.

Results and discussion

A total of 2 participants failed to rate all eight cues of the booklet, so their data were discarded from the analyses.

Figure 2 presents the mean causal ratings obtained in the testing phase of Experiment 2. As can be seen, the mean causal rating assigned to the novel compound AB was substantially greater than that given to its trained components (A and B) and to the trained control compound (EF). The reliability of this result was confirmed by a repeated measures ANOVA. There was a reliable main effect of cue, $F(7, 147) = 130.623$, $p < .001$, partial $\eta^2 = .861$, following which LSD post hoc comparisons indicated reliable differences between the novel "positive" compound AB and each of A, $t(21) = -3.727$, $p = .001$, partial $\eta^2 = .398$, B, $t(21) = -4.904$, $p < .001$, partial $\eta^2 = .534$, and EF, $t(21) = -4.519$, $p < .001$, partial $\eta^2 = .493$, and no significant differences between the novel "negative compound", CD, and C, $t(21) = -1.283$, $p = .213$, partial $\eta^2 = .073$, D, $t(21) = -1.033$, $p = .313$, partial $\eta^2 = .048$, and GH, $t(21) = -0.317$, $p = .755$, partial $\eta^2 = .005$. A separate analysis showed that the response to AB was also significantly higher than the greater of the responses to A or B for each participant, $t(21) = -3.727$, $p = .001$, partial $\eta^2 = .398$.

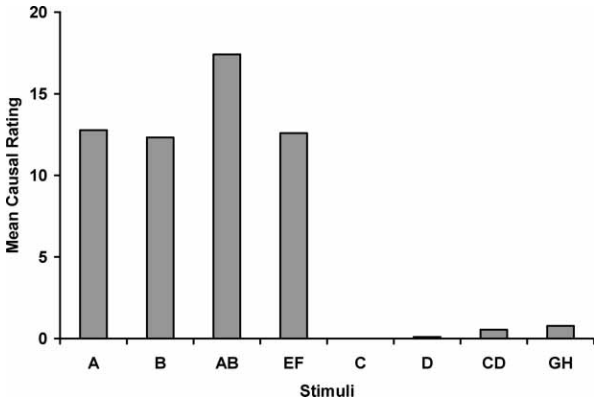

Figure 2. *Mean causal ratings to the relevant stimuli in the summation test of Experiment 2.*

These results indicate a clear summation effect despite training with categorical outcomes and testing with a likelihood scale. They are consistent with the suggestion that the failure of summation in Experiment 1 with these conditions may have been due to the participants using near-maximal values to rate the causal strength of the trained cues. Although an attempt was made to avoid bias in the testing instructions as to whether novel foods should be rated higher, lower, or equal to the trained cues, one may still be concerned about the importance of the instructions treating the training elements, A and B, differently from their novel compound, AB. Experiment 3 sought further evidence of summation using the same testing conditions as those of group qualitative/likelihood of Experiment 1, but altered training conditions to create submaximal causal ratings of the individual stimuli A and B.

EXPERIMENT 3

The outcomes of Experiments 1 and 2 suggest that an obvious feature needed to observe summation is that participants rate the trained cues A, B, and EF with submaximal causal ratings. In Experiment 3, we evaluated the summation effect using a different strategy to get submaximal scores for the A and B components. This was achieved by training A and B in compound with other cues (e.g., AC+, BD+), before evaluating the responses to A, B, and AB. It could be expected, with this training, that only part of the causal strength acquired by each compound would be carried by the A and B components, thus leading to submaximal ratings during the summation test. Evidence of summation would be like that in the previous studies, if the participants rate the novel compound AB with more causal strength than its elements, A and B. Such summation is predicted by elemental models of associative learning, such as that of Rescorla and Wagner (1972), but not by configural models, such as that of Pearce (1994), which predict that the causal ratings given to AB should be identical to that given to its elements.

The full details of the design of Experiment 3 are shown in Table 2. It included training with two compound cues, AC+ and BD+, and testing with the trained compounds, AC and BD, the single cues, A, B, C, and D, and the novel compounds, AB, AD, BC, and CD. Thus summation was evaluated in terms of the response to each of the novel compounds in comparison to their separate components. The feedback used in training and the scale used in testing were identical to those of group qualitative/likelihood of Experiment 1.

Table 2. *Design of Experiment 3*

	Training	*Test*
Summation test	AC+ BD+	AC BD
Fillers	E+ FG− H−	A B C D AB AD BC CD

Note: Letters A–H represent different foods that could be followed (+) or not followed (−) by an allergic reaction in a hypothetical patient.

Method

Participants

A total of 25 undergraduate students at the University of Talca participated in the experiment for course credit. They were tested individually and had no previous experience in similar research.

Instruments

The instruments were the same as those of Experiment 1.

Procedure

The training instructions and the structure of each trial were essentially the same as those for group qualitative/likelihood of Experiment 1, except that the number of presentations of each trial type in training was increased to 16 and that the scale used to rate the cues in testing ranged from 0 (eating the meal is very unlikely to cause an allergic reaction in Mr. X) to 10 (eating the meal is very likely to cause an allergic reaction in Mr. X.), instead of 0 to 20. The foods assigned to A–H were chocolate, cheese, mushrooms, honey, eggs, sardines, avocado, and garlic (counterbalanced).

Results and discussion

Figure 3 shows the mean causal ratings for the trained compounds (AC, BD), their component stimuli (A, B, C, D), and the novel test compounds (AB, AD, BC, and CD). As expected, the ratings for the trained compounds were higher than those for the individual stimuli, indicating that the manipulation was successful in achieving submaximal scores for the single cues. It can be also seen that the mean causal rating to the novel compounds was substantially higher

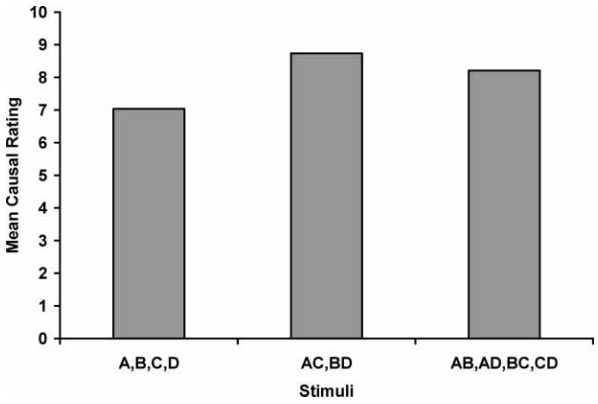

Figure 3. *Mean causal ratings to the relevant stimuli in the summation test of Experiment 3.*

than that given to the single elements, consistent with the view that the participants summated the causal strength of the constituent components when they evaluated the novel compounds. It is notable that the response to the novel compounds was somewhat less than that to the training compounds.

The reliability of these observations was evaluated by a repeated measures ANOVA. There was a reliable overall difference among the three groups of test stimuli, $F(2, 48) = 11.167$, $p < .001$, partial $\eta^2 = .318$. Separate LSD post hoc tests indicated that the mean causal rating for the individual cues was reliably lower than that for the novel, $t(24) = -2.905$, $p = .008$, partial $\eta^2 = .260$, and the trained compounds, $t(24) = -4.095$, $p < .001$, partial $\eta^2 = .411$. The fact that the participants rated the novel compound's causal strength higher than its elements is in agreement with the elemental interpretation of associative learning and inconsistent with Pearce's configural interpretation. According to Pearce's generalization rule, there should be the same response to AB, AD, BD, and CD as to A, B, C, and D.

The post hoc test indicated that the lower rating of the novel than of the trained compounds approached statistical significance, $t(24) = -1.973$, $p = .060$, partial $\eta^2 = .140$. This difference is predicted by Pearce's theory, but also by the Rescorla–Wagner Model with the assumption that there are some unique configural elements in the training compounds, the associative strength of which are not present in the novel compounds (Wagner & Rescorla, 1972).

The results from this experiment provide evidence of a reliable summation effect, using a binary outcome and a likelihood scale, under conditions not involving the special testing instructions used during Experiment 2. Although this is the first demonstration in causal learning of a summation effect from stimuli trained as part of compounds independently paired with the outcome, such a result has previously been demonstrated in Pavlovian conditioning (Rescorla, 2003).

GENERAL DISCUSSION

The three experiments reported here provide evidence of summation in human causal learning. If two stimuli are consistently paired with an outcome either individually (A+, B+; Experiments 1 and 2) or in compound with other cues (AC+, BD+; Experiment 3), the perceived causal strength assigned to the novel compound AB appears to be an additive function of the perceived causal strength of the separate elements. The results of the three experiments suggest that summation is a relatively general finding in human causal learning, independent of the specific nature of the outcome used during training (magnitude or binary) and of the kind of scale used to measure causal ratings during test (magnitude or likelihood of outcome). The observance of summation in human causal learning is not peculiar to training and testing with different magnitude outcomes, as Collins and Shanks (2006) have suggested, but can be obtained with binary outcomes in training and estimations of the likelihood of the consequence in testing. What does appear to be important is that the training and testing conditions allow the participants to give submaximal causal ratings to the elemental stimuli—that is, to avoid producing such high ratings to the elemental stimuli alone that summation is precluded by a ceiling effect.

These results are most easily explained by elemental theories of associative learning (e.g., Mackintosh, 1975; Rescorla & Wagner, 1972;

Wagner, 1981), which suppose that the associative strength of a compound is the sum of the associative strengths acquired by each of its components. They would generally be considered to be at odds with configural learning theories (Estes, 1994; Pearce, 1987), which argue that the associative strength of a compound should be the same or lower than that acquired by its individual elements. As compelling as this reasoning may be, it is important to recognize that both current elemental and configural theories have been modified in ways that make the fact of summation alone less than theoretically decisive.

Pearce (2002) has proposed that the fact of summation may be due to the contribution of contextual cues that influence the responding to both the elements and their novel compound in cases of summation. For example, if the experimental context is processed as a part of the total configuration that is associated with an outcome, then training with A+ and B+ is equivalent to training with AX+ and BX+, where X stands for the context. When A and B are compounded during test, the effective test stimulus should be regarded as ABX. By application of Pearce's generalization rule, ABX is calculated to receive 2/3 of the associative strength of AX and 2/3 of the associative strength of BX, for a total of 4/3, a value higher than that reached by AX or BX. By this reasoning Pearce's configural approach can accommodate summation effects under some circumstances by appealing to relatively salient contextual cues. It is conceivable that studies of Pavlovian conditioning where summation has been observed could involve such contextual cues (e.g., Whitlow & Wagner, 1972).[1]

In a contrasting elemental stratagem, Wagner (2003) has proposed that less than full summation may result from a process whereby stimuli, when compounded, experience a replacement of their representational elements. According to the replaced elements model (REM) the elemental representation of a compound is assumed not only to involve the addition of configural elements, but the complementary inhibition (replacement) of elements otherwise representing the compounded stimuli in isolation. Specifically, this model proposes that any stimulus is represented by a set of elements, some of which are context independent and others of which are context dependent. The context-independent elements are assumed to be activated whenever the stimulus that they represent is presented, independently of the presence or absence of any other stimulus. In contrast, the activation of context-dependent elements, although occasioned by the stimulus that they represent, depends also on the presence or absence of other "contextual" circumstances. For example, if Stimulus A alone is represented by $(a_i + a_{\sim b})$, and B alone is represented by $(b_i + b_{\sim a})$, the compound AB is represented by $(a_i + a_b + b_i + b_a)$. By this reasoning, the phenomenon of summation may fail to occur if the proportion of elements representing A and B that is context dependent, and is replaced in the compound AB, is large compared to the proportion that is context independent. Wagner (2003) has argued that this is a reasonable assumption in instances of Pavlovian conditioning with stimuli within the same modality (Aydin & Pearce, 1994), where summation has not been obtained.

The testing of these theories invites converging observations. For example, Wagner and Vogel (2008) have pointed to the usefulness of concurrent testing of summation, where observations may be in obvious line with elemental theory, as in the present experiments, and reversal learning, where the findings may seem more congruent with a configural account (Shanks, Charles, Darby, & Azmi, 1998a; Shanks, Darby, & Charles, 1998b; Williams, Sagness, & McPhee,

[1] Pearce's proposal of contextual cues in compound with A, B, and AB predicts summation by allowing for enhanced generalization between AX or BX and ABX in testing, which is not fully offset by the greater generalization between AX and BX during training. Kinder and Lachnit (2003, see also Pearce, Esber, George, & Haselgrove, 2008) have suggested a more general parametric variation in Pearce's similarity function that is in this spirit, but has not been articulated to address the specific requirements that A and B be effectively more similar to AB but not to each other.

1994; Williams, 1995). In the latter class of experiments, some cues are paired either alone or in compound with particular outcomes, and subsequently their predictive relationship with the outcome is reversed upon the introduction of novel combinations. For example, training on A+, B+, C−, D− might be followed by training on AB−, CD+, or training on A+, AB− might be followed by training on B+. A common result is that learning the second relationship does not interfere with performance to the cues in the first discrimination, as much as might be expected from an elemental account, and suggests that the cues are processed as qualitatively different stimuli (configurations) when presented alone versus in compound.

In fact, persistence in discrimination following reversal training, like the summation effect, is approachable by either Wagner's (2003) REM or Pearce's (2002) configural theory. In the case of REM, it needs to be assumed that the proportion of replaced elements is large, just as necessarily in the case of Pearce's configural theory it needs to be assumed that the salience of the contextual cues is small. That is, the assumptions made by each of the two theories to generate persistence in discrimination are just the opposite in direction to those required to generate summation. The critical theoretical test is not whether summation occurs, or whether persistence in discrimination occurs, but, rather, whether, with a given assumption about the degree of replacement or the salience of the contextual cues, one or the other model can accommodate both the degree of summation and the degree of persistence in discrimination that is empirically observed (see Wagner & Vogel, 2008, for quantitative simulations that illustrate the possibilities).

It remains to be seen how well either elemental or configural theories can deal satisfactorily with the converging results from these and other different tests in human causal learning. A challenging possibility that has been proposed (Melcher, Shanks, & Lachnit, 2008; Williams, 1995) is that stimulus processing involves both elemental and configural coding (e.g., Fanselow, 1999; Pearce & Bouton, 2001; Wagner, 2003) and that the predominance of one or another is a function not only of the task demands, but of the individuals' learning history. Consistent with this notion, there is some evidence that training with elemental discriminations disposes participants to use elemental coding in subsequent tasks, whereas training with configural discriminations disposes participants to use configural coding in subsequent tasks (Williams & Braker, 1999; Williams et al., 1994). It is not known how general is the influence observed across different stimuli or tasks, but a major challenge to our theories becomes how to accommodate whatever flexibility of processing exists. From a Pearcian perspective, it could be via differential attention to the contextual cues. From a replaced-elements perspective it could be via differential attention to context-independent versus context-dependent elements.

Original manuscript received 13 March 2008
Accepted revision received 21 July 2008
First published online 1 December 2008

REFERENCES

Aydin, A., & Pearce, J. M. (1994). Prototype effects in categorization by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*, 264–277.

Aydin, A., & Pearce, J. M. (1995). Summation in autoshaping with short- and long-duration stimuli. *Quarterly Journal of Experimental Psychology*, *42B*, 215–234.

Aydin, A., & Pearce, J. M. (1997). Some determinants of response summation. *Animal Learning and Behavior*, *25*, 108–121.

Collins, D. J., & Shanks, D. R. (2006). Summation in causal learning: Elemental processing or configural generalization? *Quarterly Journal of Experimental Psychology*, *59*, 1524–1534.

Estes, W. K. (1994). *Classification and cognition*. Oxford, UK: Oxford University Press.

Fanselow, M. S. (1999). Learning theory and neuropsychology: Configuring their disparate elements in the hippocampus. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 275–283.

Kehoe, E. J., Horne, A. J., Horne, P. S., & Macrae, M. (1994). Summation and configuration between and within sensory modalities in classical conditioning

of the rabbit. *Animal Learning and Behavior*, *22*, 19–26.

Kinder, A., & Lachnit, H. (2003). Similarity and discrimination in human Pavlovian conditioning. *Psychophysiology*, *40*, 226–234.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

Melchers, K. G., Shanks, D. R., & Lachnit, H. (2008). Stimulus coding in human associative learning: Flexible representations of parts and wholes. *Behavioural Processes*, *77*, 413–427.

Myers, K. M., Vogel, E. H., Shin, J., & Wagner, A. R. (2001): A comparison of the Rescorla–Wagner and Pearce models in a negative patterning and a summation problem. *Animal Learning and Behavior*, *29*, 36–45.

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

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607.

Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning and Behavior*, *30*, 73–95.

Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, *52*, 111–139.

Pearce, J. M., Esber, G. R., George, D. N., & Haselgrove, M. (2008). The nature of discrimination learning in pigeons. *Learning and Behavior*, *36*, 188–199.

Pearce, J. M., George, D. N., & Aydin, A. (2002). Summation: Further assessment of a configural theory. *Quarterly Journal of Experimental Psychology*, *55B*, 61–73.

Rescorla, R. A. (1997). Summation: Assessment of a configural theory. *Animal Learning and Behavior*, *25*, 200–209.

Rescorla, R. A. (2003). Elemental and configural encoding of the conditioned stimulus. *Quarterly Journal of Experimental Psychology*, *56B*, 161–176.

Rescorla, R. A., & Coldwell, S. E. (1995). Summation in autoshaping. *Animal Learning and Behavior*, *23*, 314–326.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non reinforcement. 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., Charles, D., Darby, R. J., & Azmi, A. (1998a). Configural processes in human associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1353–1378.

Shanks, D. R., Darby, R. J., & Charles, D. (1998b). Resistance to interference in human associative learning: Evidence of configural processing. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 136–150.

van Osselaer, S. M. J., Janiszewski, C., & Cunha, M. (2004). Stimulus generalization in two associative learning processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 626–638.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wagner, A. R. (2003). Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology*, *56B*, 7–29.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In M. S. Halliday & R. A. Boakes (Eds.), *Inhibition and learning* (pp. 301–336). San Diego, CA: Academic Press.

Wagner, A. R., & Vogel, E. H. (2008). Configural and elemental processing in associative learning. *Behavioural Processes*, *77*, 446–450.

Whitlow, J. W., & Wagner, A. R. (1972). Negative patterning in classical conditioning—summation of response tendencies to isolable and configural components. *Psychonomic Science*, *27*, 299–301.

Williams, D. A. (1995): Forms of inhibition in animal and human learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 129–142.

Williams, D. A., & Braker, D. S. (1999). Influence of past experience on the coding of compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 461–474.

Williams, D. A., Sagness, K. E., & McPhee, J. E. (1994). Configural and elemental strategies in predictive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 694–709.