The Quarterly Journal of Experimental Psychology Section B

ISSN: 0272-4995 (Print) 1464-1321 (Online) Journal homepage: http://www.tandfonline.com/loi/pqjb20

Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity

Chris J. Mitchell & Peter F. Lovibond

To cite this article: Chris J. Mitchell & Peter F. Lovibond (2002) Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity, The Quarterly Journal of Experimental Psychology Section B, 55:4, 311-329

To link to this article: http://dx.doi.org/10.1080/02724990244000025

Published online: 24 Sep 2010.

Submit your article to this journal

Article views: 269

View related articles

Citing articles: 23 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=pqjb20

Download by: [The UC Irvine Libraries] **Date:** 16 December 2015, At: 16:30

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2002, *55B* (4), 311–329

Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity

Chris J. Mitchell and Peter F. Lovibond
University of New South Wales, Sydney, Australia

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Blocking was observed in two human Pavlovian conditioning studies in which colour cues signalled shock. Both forward (Experiment 1) and backward (Experiment 2) blocking was demonstrated, but only when prior verbal and written instructions suggested that if two signals of shock (A+ and B+) were presented together, a double shock would result (AB++). In this case, participants could assume that the outcome magnitude was additive. Participants given non-additivity instructions (A+ and B+ combined would result in the same outcome, a single shock) failed to show blocking. Modifications required for associative models of learning, and normative statistical accounts of causal induction, to account for the impact of additivity instructions on the blocking effect, are discussed. It is argued that the blocking shown in the present experiments resulted from the operation, not of an error-correction learning rule, nor of a simple contingency detection mechanism, but of a more complex inferential process based on propositional knowledge. Consistent with the present data, blocking is a logical outcome of an A+/AB+ design only if participants can assume that outcomes will be additive.

In a blocking procedure, a conditioned stimulus (CS), A, is first paired with an unconditioned stimulus (US), such as electric shock, and subsequently a compound AB is followed by shock. The common result is that conditioning to B is attenuated as a result of pre-training with A (Kamin, 1969). The blocking effect, and other examples of selective learning, have been demonstrated across a wide variety of procedures and species, and they are central to contemporary animal learning theory (Mackintosh, 1975; Miller & Schachtman, 1985; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Sutherland & Mackintosh, 1971; Wagner, 1981). These theories explain conditioned responses (CRs) as resulting from CS-induced activation of the US node, by way of an associative link. The inability of the CS to excite activation at the US node following a blocking procedure is usually attributed to a failure on compound trials to process either the CS (Mackintosh, 1975; Pearce & Hall, 1980; Sutherland & Mackintosh, 1971) or the US (Rescorla & Wagner, 1972; Wagner, 1981). In contrast, comparator models

Requests for reprints should be sent to Chris Mitchell, School of Psychology, University of New South Wales, Sydney NSW 2052, Australia. Email: chris.mitchell@unsw.edu.au

This work was supported by a grant from the Australian Research Council.

© 2002 The Experimental Psychology Society
http://www.tandf.co.uk/journals/pp/02724995.html DOI:10.1080/02724990244000025

see blocking as a performance effect, and attenuated responding to B resulting from the relatively high associative value of the comparator cue A, with which B was trained (Miller & Schachtman, 1985).

More recently, the predictions of these associative models have been applied to human learning with considerable success. For instance, studies of human causality judgement have shown evidence for blocking (Chapman & Robbins, 1990; Dickinson, Shanks, & Evenden, 1984), overshadowing (Price & Yates, 1993), conditioned inhibition (Chapman & Robbins, 1990), and contingency effects (Dickinson et al., 1984) similar to those found in the animal conditioning literature. Thus, models of animal associative learning may also provide a good account of human causal judgement (Dickinson et al., 1984).

However, the blocking effects observed in these causal judgement tasks have often been weak, and complete blocking is typically not observed. In addition, there is only sparse evidence of blocking effects in human Pavlovian conditioning, using procedures that more closely resemble those used in the animal conditioning literature. Pellon and Montano (1990), Pellon, Montano, and Sanchez (1995), Kimmel and Bevill (1991, 1996), and Hinchy, Lovibond, and Ter-Horst (1995) have provided some evidence for weak blocking effects in human autonomic conditioning, whereas Davey and Singh (1988) and Lovibond, Siddle, and Bond (1988) report failures to show blocking in the same types of procedure across a number of experiments.

The present paper investigates the possibility that the difficulty in obtaining strong evidence for blocking in human participants, in both Pavlovian conditioning and causality judgement tasks, may have a common source. That is, if human participants solve both learning tasks through inferential processes based on propositional knowledge (as has been suggested by Lovibond, in press), rather than through the formation of associative links, a blocking effect would not be expected to occur under all conditions. It does not directly follow from the propositions "A leads to shock" (A+) and "A and B lead to shock" (AB+) that "B does not lead to shock" (B–). Given the two antecedent statements, B is ambiguous with respect to its relationship to the outcome; it may or may not lead to shock when presented alone.

It is only possible to disambiguate the causal status of B if additional assumptions about cue combination are made. For instance, it can be concluded that B is non-causal (blocking) if, in addition to the A+/AB+ information, it is known that, generally, the presentation of a compound of two causes of shock (A+ and B+) results in a shock of greater magnitude (AB++). That is, when causal cues are combined, the magnitude of the outcome is additive.[1] In this case, when presented with A+/AB+ trials, it can be concluded that B is safe (B–) because, had B been a predictor of shock itself (B+), then the AB compound would have been followed by a shock of greater intensity (AB++). The AB compound was not followed by a shock of greater intensity, thus B is not causal. According to this argument, failures to demonstrate blocking in human electrodermal conditioning preparations would be expected because the outcome is binary, and thus additivity cannot be assumed. That is, the participants are asked to set a level of shock that is uncomfortable (but not painful) at the beginning of the study, and they are told that no greater shock will be administered to them at any stage. This maximum level of shock is

[1]The term "additive" is not used here to imply that participants given two cues for an outcome will expect two shocks, or a shock of exactly double the intensity, but merely a larger outcome than that which followed either of the signals individually.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

then used throughout the study, and on each occasion that the reinforced cues are presented; the shock administered on A+ trials is the maximum possible.

The effect of assumptions of additivity on blocking were recently examined in a causality judgement task in this laboratory. Lovibond, Been, Mitchell, Bouton, and Frohardt (2001) manipulated participants' assumptions of outcome magnitude additivity by providing evidence for additivity or non-additivity in an early phase of training. Participants were asked to play the role of an allergist dealing with a hypothetical patient, Mr. X. They were exposed to a series of trials consisting of meals in which the patient ate either one or two foods. Following this training, participants rated each food for its ability to cause an allergic reaction in Mr. X. All participants received A+ and AB+ trials (blocking). For half of the participants, A+ trials preceded AB+ trials (forward blocking), whereas the remaining participants received the trials in reverse order (backward blocking). Outcome additivity pre-training was given to half of the participants in each group before the target cues were presented. This pre-training involved the presentation of E+ and F+ trials intermixed with EF++ trials (the compound resulted in a "strong allergic reaction"). The remaining participants received non-additivity training. That is, E+ and F+ trials were intermixed with EF+ trials. The results were clear. Very strong forward and backward blocking appeared following additivity training. Following non-additivity training, forward blocking was weak, and backward blocking did not appear at all. De Houwer, Beckers, and Glautier (2002) observed a similar effect in their causal judgement task; strong blocking effects were observed, but only in participants for whom outcome magnitude additivity was possible. Finally, Cheng (1997) has made a very similar argument with respect to the additivity of outcome probability. In her power PC model of causal induction, blocking should not be expected if the probability of the outcome on A+ trials is at ceiling ($p = 1$).

The focus of the present paper is whether such a propositional reasoning process is also the most persuasive account of the generation of CRs observed in human Pavlovian conditioning studies. It might be thought that conscious awareness, and therefore propositional knowledge, is not required for the generation of CRs. It is commonly argued that there exist two levels of learning, one propositional system responsible for the knowledge that is available to consciousness, and a second lower level system that gives rise to CRs in the absence of awareness (Razran, 1955; Squire, 1994; see Lovibond & Shanks, 2002, for review). The present studies assess whether assumptions of magnitude additivity influence the process of CR generation as they do causality judgement.

EXPERIMENT 1

The first study investigated the effect of participants' assumptions about outcome additivity in a traditional "forward" blocking design. Simple visual cues were paired with shock, and participants' expectancy of shock and skin conductance response (SCR) on presentation of these stimuli were recorded. All participants received the same training: A+ trials, followed by AB+ and CD+ trials. A blocking effect would be observed if responses to B on test were lower than those to C or D, the "overshadowed" control cues. Half of the participants were given verbal and written instructions indicating that outcomes would be additive, whereas the remaining participants were instructed that outcomes would be non-additive. According to a propositional reasoning model, participants given additivity instructions should show

blocking to a greater extent than those given non-additivity instructions. That is, the "blocked" cue B should generate weaker shock expectancies and SCRs relative to C. The within-subjects design, with respect to the blocking manipulation, was expected to increase the sensitivity of the test relative to previous autonomic conditioning studies that have examined blocking, but not to affect the additivity manipulation.

The use of verbal instructions in manipulating the participants' assumptions about additivity, rather than prior training as used in the causality judgement experiments of Lovibond et al. (2000), provides a further test of the levels of learning hypothesis. If the generation of propositional knowledge and CRs results from the working of two separate systems, then verbal instructions of any kind would not be expected to affect CR generation. Thus, although the levels of learning position allow that the pattern of "conscious" shock expectancies may differ across additive and non-additive groups, as they did in the causality judgement task (Lovibond et al., 2001), the additivity instructions would not be expected to influence the extent of blocking observed on the SCR measure.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Method

Participants

The participants were 32 undergraduate and postgraduate students (15 male and 17 female, with an age range of 18–28 years) at the University of New South Wales. The participants volunteered to take part in the experiment; 21 participated as part of a course requirement, and the remaining 11 participants were given $10 AUS to compensate them for their time. Participants were alternately allocated to the additive and non-additive groups.

Apparatus

An IBM-compatible computer with a Med PC control interface was used to present all stimuli and to record the shock expectancy ratings (at 500-ms intervals) and skin conductance levels (at 200-ms intervals). Participants were tested individually in a dimly lit room. Except for when instructions were being given, or the shock level was being set, participants wore headphones attached to a sound generator producing 80-dB white noise in order to mask background noise. The stimuli were coloured blocks of black, red, blue, yellow, green, and purple on a white background presented on a 30-cm colour computer monitor approximately 100 cm in front of the participant. The blocks were approximately 6 cm square and appeared in the left or right half of the screen, centred vertically. When appearing together they were separated by 3 cm. The monitor was also used to present a brief version of the instructions directly before commencement of the experiment (see Procedure).

Skin conductance was measured through electrodes attached to the distal segments of the second and third fingers of the participant's non-preferred hand, and it was digitally recorded by the Med PC interface. Shock electrodes were attached to the proximal and medial segments of the first (index) finger of the same hand. Participants used their preferred hand to record their subjective moment-to-moment expectancy of shock on a continuous, 180-degree, rotary dial attached to the arm of seat. The semicircular dial was labelled *Expectancy of shock*, with the left extreme labelled *Certain no shock*, and the right extreme *Certain shock*, with 0%, 20%, 40%, 60%, 80%, and 100% marked at 36-degree intervals.

Procedure

The general procedure followed that described in Hinchy et al. (1995). Participants were fitted with electrodes and were led through a work-up procedure to select a "definitely uncomfortable, but not

painful" shock level. It was made clear that this level of shock would not be exceeded at any point during the experiment. Participants were then taken into the experimental room and given spoken and written instructions about the purpose of the experiment and the use of the expectancy pointer. They were told that they should move the expectancy pointer as often as they wished, so that it always indicated their current degree of expectancy of shock at the end of the CS. Participants were specifically directed to work out which stimuli led to shock and which did not. They were informed that shocks would occur only at the end of a trial, as the stimulus was taken from the screen. The timing of events followed that developed by Lovibond (1992). Throughout the experiment, CS durations varied randomly between 15 s and 35 s, whereas shock durations were always 0.5 s. Inter-trial intervals varied randomly between 30 s and 50 s (mean = 40 s).

The six colours were assigned labels A to F. CSs A through D were counterbalanced in the following manner. For half of the participants, blue served as CS A (the "blocking" cue) and yellow as CS D (the "overshadowing" cue). This assignment was reversed for the remaining participants. Orthogonally, green served as CS B (the "blocked" cue, paired with A in training), whereas purple served as CS C (the "overshadowed" cue, paired with D in training) for half of the participants; the remaining participants received the reverse assignment. Lastly, red served as CS E, and black served as CS F for all participants (the two "safe" cues).

A mixed design was used in which all participants were exposed to both the blocking relationship (A+/AB+) and the overshadowing relationship (CD+), and additivity instructions were manipulated between participants. The stimuli were presented according to the schedule in Table 1. During the element phase, Cue A was presented twice and followed by shock on each occasion, whereas Cues E and F and the compound EF were presented once each with no outcome. During the compound phase, the compounds AB and CD were each presented twice, followed by shock. Intermixed were presentations of E, F (once each), and EF (twice), again with no outcome. Within each training phase, trial order was random. On test, all the cues except D were presented individually; D was not presented because it had the same status as C. Only Cue A was followed by shock and was presented first on test to all participants, whereas the presentation order of B and C was counterbalanced across participants. Lastly, presentations of each of the stimuli A, B, and C during the test phase were separated by presentations of the cues E or F (in random order). Thus, the four test orders were AEBFC, AECFB, AFBEC, and AFCEB.

TABLE 1
Design of Experiment 1

Phase					
Element		*Compound*		*Test*	
Stimulus	*No. of trials*	*Stimulus*	*No. of trials*	*Stimulus*	*No. of trials*
A+	2	AB+	2	A+	1
		CD+	2	B–	1
E–	1	E–	1	C–	1
F–	1	F–	1	E–	1
EF–	1	EF–	2	F–	1

Note: Letters A to F refer to conditioned stimuli; + and – refer to the presence and absence, respectively, of electric shock after the CS. Within each phase, trial types were intermixed, except that A+ was presented first on test to all participants.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

The between-subjects manipulation of additivity was administered by instruction. The additive group received additivity instructions, and the non-additive group received non-additivity instructions.

The additivity instructions were as follows:

> The last thing that you need to know is that you may receive a double shock on some trials. For example, when two colours that individually lead to shock are presented together, the pair will be followed by a double shock. That is, one shock followed by a second shock. So, if pink alone is followed by shock, and orange alone is followed by shock, then it would follow that if pink and orange are shown together, they will be followed by a double shock.

The non-additivity instructions were as follows:

> The last thing you need to know is that the relationship between the colours and shock is consistent across trials. For example, if two colours that individually lead to shock are presented together, they will also be followed by shock. So, if pink alone is followed by shock, and orange alone is followed by shock, then it would follow that if pink and orange are shown together, they will also be followed by shock.

Thus, following instructions, participants in the additive group, when presented with a compound stimulus made up of two predictors of shock, should expect two shocks. Conversely, participants in the non-additive group should expect a single shock if presented with the same two stimuli. Of course, no participant actually received a double shock at any time during the experiment; the between-subjects variable concerned the manipulation of participants' beliefs as to whether double shocks might occur and, if so, under what conditions. These instructions were presented on the screen for 40 s, along with a summary of the previous instructions that participants would be presented with a series of colours, that they should try to work out which colours lead to shock, and on the use of the expectancy dial. The instructions were then removed and, following a 20-s delay, the stimulus presentation schedule was initiated.

Participants were classified as being aware of the stimulus relations if their mean expectancy rating on the final A+ trials was at least 50 points higher than their mean expectancy rating on the last E and F trials. The value of 50 points was chosen as a relatively conservative criterion so as to reduce the possibility of misclassification.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Scoring and analysis

The electrodermal measure taken was change in tonic skin conductance level (SCL), calculated for each trial as the difference between the mean SCL during the CS (excluding the first 10 s) and the mean SCL for the 10-s baseline period immediately prior to CS onset (Lovibond, 1992). In order to eliminate the large individual differences between participants in the magnitude of their SCL change scores, these data were mean-corrected. This procedure, described in detail in Lovibond (1992), expresses the SCL change score on each trial as a proportion of the mean SCL change score for that participant over all trials. Thus a score of 1 indicates a typical response for that participant, 0 indicates no response, and a score above 1 indicates a higher than typical response. Expectancy ratings were analysed in their raw form as percentages. A set of orthogonal contrasts was tested using a multivariate, repeated measures model (O'Brien & Kaiser, 1985) for both the SCL and US expectancy measures. A significance level of $p < .05$ was set for all of the following statistical analyses.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Results and discussion

Three participants were rejected from the analysis: one participant was classed as unaware, one failed to use the expectancy dial at any stage during the study, and one participant revealed that he was colour-blind at the end of testing. These three were replaced with further participants in order to equate the number of participants in each counterbalancing cell (see Procedure).

Expectancy ratings

Expectancy data are shown in Figure 1 for the additive group (top panel) and the non-additive group (bottom panel). Participants in both groups had strong expectancies of shock on the second A+ trial (at the end of the element phase), relative to both the E– and the F– trials (elements not followed by shock). The contrast comparing the second A+ trial with an average of E– and F– was reliable, $F(1, 30) = 423.1$. In addition, there was no interaction between this contrast and the between-subjects factor of group ($F < 1$), suggesting that Phase 1 was equivalently effective across groups. Also, strong shock expectancy was found in both groups on the second AB+ and CD+ trials in the compound phase, as compared to the second compound EF– trial. A contrast comparing shock expectancy on the second EF– trial with an average of the second AB+ and CD+ trials was highly reliable, $F(1, 30) = 1085.8$, and again,

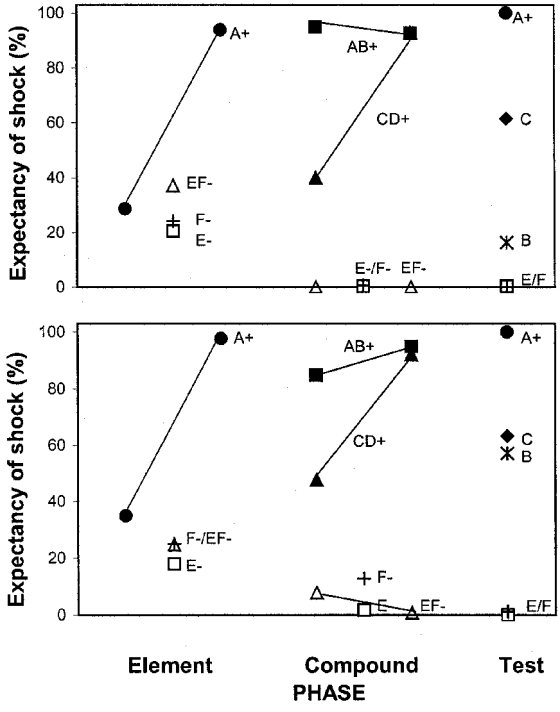

Figure 1. Mean expectancy ratings across the three phases of Experiment 1. The top panel shows responses for the additive group; the bottom panel shows responses for the non-additive group.

did not interact with the grouping factor ($F < 1$). AB+ and CD+ trials did not differ from one another ($F < 1$), and this contrast did not interact with the grouping factor ($F < 1$). Thus, the element and compound phases resulted in a strong expectancy of shock on the final A+ trial, and on the final AB+ and CD+ trials, which did not differ across groups.

The critical comparison, however, is that between the "blocked" and "overshadowed" elements B and C, respectively, on test. It would appear that participants' expectation of shock on the B– trial was lower than that on the C– trial on test, but only in the additive group. The contrast comparing shock expectancies on B– and C– trials was significant, $F(1, 30) = 16.7$, as was the interaction between this contrast and the grouping factor, $F(1, 30) = 9.7$. Two post hoc tests were conducted in order to identify the source of this interaction. A difference between shock expectancies on B– and C– trials was found in the additive group, $F(1, 15) = 54.3$, but not in the non-additive group, $F < 1$. This would suggest that blocking occurred, but only when additive instructions were given to the participants.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Skin conductance

Figure 2 shows the SCL data for both the additive group (top panel) and the non-additive group (bottom panel). The SCL data for the element and compound phases appear very similar to the expectancy data for the same phases. Responding was higher on A+ than on E– and F– trials at the end of element training. A contrast comparing the second A+ trial with an average of the E– and F– trials was found to be reliable, $F(1, 30) = 20.5$, and this contrast did not

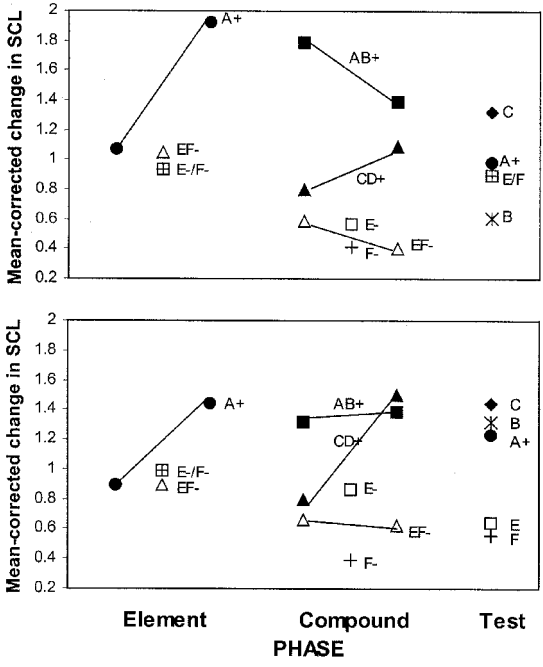

Figure 2. Mean SCR scores across the three phases of Experiment 1. The top panel shows responses for the additive group; the bottom panel shows responses for the non-additive group.

interact with groups, $F(1, 30) = 2.9$, *ns*. Also, responding to the second presentation of the compounds AB+ and CD+ was higher than that to the compound EF–, $F(1, 30) = 55.6$, and, again, no interaction with groups was found, $F < 1$. AB+ and CD+ trials did not differ from one another, $F < 1$, and this contrast did not interact with the grouping factor, $F(1, 30) = 1.4$, *ns*. Thus, conditioning to the element A+ and to the compounds AB+ and CD+ was reliable and did not differ across groups. On test, responding to A+ was higher than the average of that to E– and F–, $F(1, 30) = 5.3$, and no interaction with groups was found, $F(1, 3 0) = 3.1$, *ns*.

There was no evidence of overshadowing, with SCRs to A being lower than those to B and C in most cases. However, SCRs are generally higher to novel stimuli (for instance a habituation effect can be observed across presentations of E and F). Thus, the reactions to B and C are not comparable to those to A; B and C, but not A, are presented alone for the first time on test and will therefore generate CRs in part through their novelty. The novelty of Cues B and C is equal, and therefore the more important comparison between these two can be made. The data concerning responding to B and C on test are similar to those observed on the expectancy measure. That is, B and C would appear to differ in the additive group but not in the non-additive group. A contrast revealed an overall difference between B and C across the two groups, $F(1, 30) = 9.2$, and this contrast interacted with the grouping factor, $F(1, 30) = 4.705$. When further post hoc analyses were conducted, a difference between B and C was found in the additive group, $F(1, 15) = 16.3$, but not in the non-additive group ($F < 1$). Thus, just as for the expectancy data, blocking was demonstrated, but only when participants were given additivity instructions.

The participants given additivity instructions showed convincing blocking on both expectancy and SCR measures, an effect that was completely absent following nonadditivity instructions. This finding argues against the levels of learning hypothesis; it would appear that the propositional system, which, according to the levels of learning hypothesis, is separate from the non-propositional learning system, is nevertheless able to affect autonomic CRs. The effect of additivity instructions adds to the wealth of data indicating that "conditioned" responses can be induced by instruction, or a combination of instruction and experience (see Lovibond & Shanks, 2002, for review).

From the perspective of attempts to extend contemporary animal learning theory to the explanation of human cognition, it is reassuring that the central phenomenon of animal conditioning—blocking—can be demonstrated in human participants. However, the finding that participants must assume outcome additivity in order for blocking to be observed presents difficulties for models that explain the phenomenon as the reduction of processing of an expected or unsurprising cue (Karnin, 1969; Rescorla & Wagner, 1972). In both the additive and non-additive groups, the shock was expected on AB+ trials as a result of A+ pre-training. The single shock following the AB compound is no more expected (less surprising) for participants in the additive group than it is for participants in the non-additive group. The present data are, however, consistent with any model that suggests that human participants, when presented with the task of determining which of a range of cues predicts the outcome, use some propositional reasoning process based on a combination of information acquired from training trials and from any instructions given.

According to a propositional account, whether A+ trials precede or follow AB+ trials (forward and backward blocking, respectively) ought to be of no consequence, assuming that all information is available on which to base any inference. Whereas forward blocking such as that

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

demonstrated by Kamin (1969) is common in the animal learning literature, reports of backward blocking, an attenuation of responding to B as a result of A+ trials following AB+ trials, are rare. It would have to be supposed that, in backward blocking, the associative strength of B is changed retrospectively, either during A+ or on test (e.g., Dickinson & Burke, 1996). Backward blocking is not an uncommon finding in human causal judgement experiments (e.g., Shanks, 1985; Wasserman & Berglan, 1998), although Lovibond et al. (2001) found that the additivity training that they used strongly determined the presence of backward blocking. A demonstration of backward blocking in human conditioning would add to the parallels between causal judgement and conditioning. Furthermore, if causal judgements and autonomic CRs result from similar processes, then an effect of additivity in a backward blocking design, as was found by Lovibond et al. (2001), would be expected.

It has been suggested recently that the difficulty in demonstrating backward blocking in animals lies in the task used rather than the species tested, and specifically in the nature of the outcome (Miller & Matute, 1996). They demonstrated backward blocking in animals, but only when the outcome was biologically non-significant, and argued that when biologically significant outcomes are used, backward blocking will not be observed. It is clear that the outcomes used in human causality judgements, for instance an allergy suffered by a fictitious character, are similarly low in biological significance. It is possible then, that if the biological significance of the outcome were high in a human learning preparation, backward blocking might be difficult to detect. An electric shock such as those used in the present experiments might be considered to be biologically significant and thus provide a test of this hypothesis. A demonstration of backward blocking in human Pavlovian conditioning would suggest that the distinction between the human and animal data with respect to the phenomenon of backward blocking is not in the nature of the outcome. As a result, it would have to be accepted that humans and non-human animals differ significantly in their mental processes, either qualitatively or quantitatively. Experiment 2 tested this notion.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

EXPERIMENT 2

The present study was conducted as an attempt to obtain evidence of backward blocking in human Pavlovian conditioning while manipulating assumptions of additivity. A demonstration that additivity instructions enhance backward blocking in human Pavlovian conditioning would add to the existing evidence from Experiment 1 that blocking in human conditioning results from a propositional reasoning process rather than from a failure of the blocked cue to gain associative strength.

Method

The method was the same as that used in Experiment 1 except in the following respects.

Participants

The participants were 40 undergraduate and postgraduate students (16 male and 24 female, with an age range of 18–33 years) at the University of New South Wales. The participants volunteered to take part in the experiment; 33 participated as part of a course requirement, and the remaining 7 participants were given $10 AUS to compensate them for their time.

Apparatus and procedure

The apparatus and procedure were identical to those used in Experiment 1 except that the element and compound phases were reversed.

Results and discussion

The scoring, analyses, and criteria for participants' rejection were the same as those used in Experiment 1. Only one participant was classified as unaware and was replaced by a further participant.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Expectancy ratings

The expectancy data are presented in Figure 3 for both the additive group (top panel) and the non-additive group (bottom panel). As in Experiment 1, expectancy of shock on AB+ and CD+ trials was high at the end of the compound phase compared to that on the second EF– trial. A contrast comparing both the second AB+ and the second CD+ trial with the second EF– trial was significant, $F(1, 38) = 883.5$. This contrast did not interact with the grouping factor ($F < 1$). The second AB+ and CD+ trials did not differ from one another ($F < 1$), and this contrast did not interact with the grouping factor ($F < 1$). A strong expectancy of shock to the A+ element was also seen at the end of element training. The second A+ trial was compared to an average of E– and F–, and the difference was found to be reliable, $F(1, 38) = 990.2$.

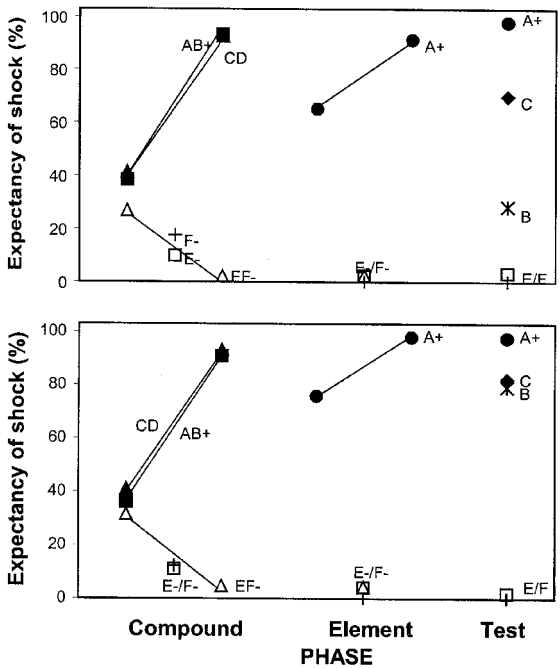

Figure 3. Mean expectancy ratings across the three phases of Experiment 2. The top panel shows responses for the additive group; the bottom panel shows responses for the non-additive group.

Again, the contrast did not interact with groups ($F < 1$). Thus, the compound and element phases were successful in establishing an expectancy of shock to the AB and CD compounds, and to the A element. The absence of any interaction between these contrasts and the grouping factor indicates that conditioning was equally strong in both groups.

More important, however, as was observed in Experiment 1, expectancy of shock on the C– trial on test was higher than that on the B– trial, but only in the additive group. The contrast comparing B and C was found to be significant, $F(1, 38) = 26.0$, and interacted reliably with the grouping factor, $F(1, 38) = 19.4$. Post hoc analyses revealed a difference in shock expectancy between B– and C– trials in the additive group, $F(1, 19) = 27.8$, but not in the non-additive group ($F < 1$), suggesting that blocking occurred only when participants were given additivity instructions.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

Skin conductance

The SCL data are presented in Figure 4 for the additive group (top panel) and the non-additive group (bottom panel). Just as in Experiment 1, the SCL data are similar in pattern to the expectancy data. Thus, greater mean responding to AB+ and CD+ than to the EF– compound was observed at the end of the compound phase; a contrast comparing the second AB+ and CD+ trials with the EF– trial was reliable, $F(1, 38) = 9.7$, and this contrast did not interact with groups, $F(1, 38) = 1.2$. In addition, the second AB+ and CD+ trials did not differ from one another ($F < 1$), and this contrast did not interact with the grouping factor ($F < 1$). Also,

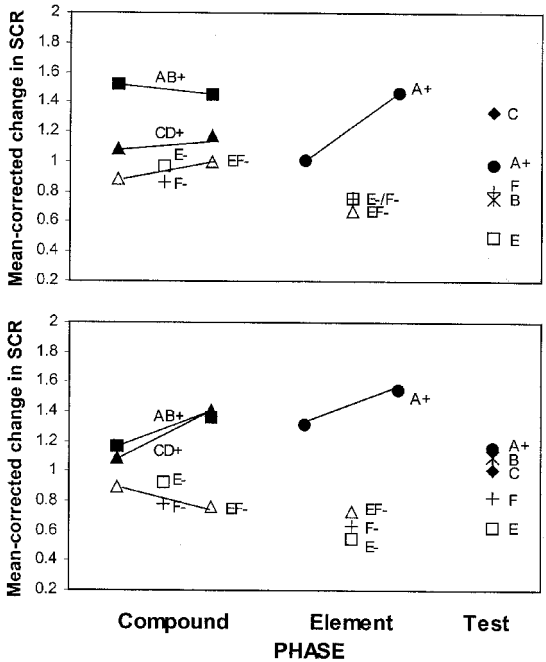

Figure 4. Mean SCR scores across the three phases of Experiment 2. The top panel shows responses for the additive group; the bottom panel shows responses for the non-additive group.

responding on the second A+ trial of the element phase was higher than that on an average of E– and F– trials, $F(1, 38) = 28.2$, and this contrast also did not interact with groups ($F < 1$). Finally, on test, the A+ trial led to greater responding than did the average of E– and F– trials, $F(1, 38) = 6.1$, and again, this contrast did not interact with the grouping factor ($F < 1$). Therefore, responding was successfully conditioned to the AB and CD compounds and to the A element, and this conditioning was equivalent across the two groups.

Before turning to the test data, one final aspect of the training data deserves mention. The very first AB+ and CD+ trials in the compound training phase appear to show greater responding than the first EF– trial. Responding to AB+ and CD+ combined was in fact reliably greater than that to EF– on both the expectancy measure, $F(1, 38) = 7.341$, and the skin conductance measure, $F(1, 38) = 4.86$. This effect occurred despite counterbalancing of trial order. One possible explanation is that participants applied the gamblers fallacy; if they received shock on the first trial, they thought that shock on the second trial was less likely and vice versa. Alternatively, the high level of the SCR on the first AB and CD trials might be attributed to the fact that arousal generally drops after the first shock has been received. Thus, if participants receive a shock on their first trial (e.g., AB+), then their responding will be lower on the following trial, whereas if the first trial is safe (EF–), then arousal will remain high on the second trial. Overall, responding to the first shocked cues will be higher than those to the first safe cues (see Lovibond et al., 1988, for a similar effect).

On test, as in Experiment 1, responding appeared to be higher to C than to B in the additive group but not in the non-additive group. The contrast comparing B and C averaged across groups was found to be unreliable, $F(1, 38) = 3.5$, *ns*, but the interaction between this contrast and the grouping factor was significant, $F(1, 38) = 6.2$. Post hoc analyses revealed a difference between B and C in the additive group, $F(1, 19) = 7.6$, but not in the non-additive group ($F < 1$). Again, as was the case for the expectancy data in the present experiment, and for both SCL and expectancy data in Experiment 1, a blocking effect was observed in the additive group, but not in the non-additive group.

There was one unexpected aspect to the pattern of data on test, that Cue C showed a higher SCR in the additive group than in the non-additive group. The propositional account suggests the opposite pattern; as both C *and* D might be causal in the non-additive group, but only C *or* D in the additive group, responding to C ought to be higher than that to A, E, and F in the former group. Post hoc contrasts were conducted comparing C to A and to E/F. Neither revealed a reliable difference across the two groups with respect to C. There was no overall difference between C and A ($F < 1$), and this contrast did not vary across groups, $F(1, 38) = 3.17$. Lastly, although C differed from the average of E and F, $F(1, 38) = 10.97$, this difference was equivalent in the two groups, $F(1, 38) = 1.86$. Thus, although the pattern of responses to C is interesting, and not what would be expected according to the propositional account, there is no reliable evidence that participants showed greater arousal to C following additive instructions.

The present results represent the first evidence for backward blocking in the human Pavlovian conditioning literature. There was little indication of a difference between backward blocking and the forward blocking effect found in Experiment 1. As in Experiment 1, the additivity assumption had a dramatic effect on the outcome; very strong blocking was observed on both expectancy and SCR measures in the additive group, but no evidence of blocking was seen on either measure in the non-additive group. The expectancy data are

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

entirely consistent with what one would expect from the demonstrations of backward blocking in causality judgement tasks (Lovibond et al., 2001). The SCR data are further evidence that information in propositional form (the additivity instructions) can interact with that gained through direct experience (the training trials) to generate conditioned responses such as an increase in autonomic arousal.

GENERAL DISCUSSION

Across the two experiments convincing evidence of forward (Experiment 1) and backward (Experiment 2) blocking in human Pavlovian conditioning with shock as the US was observed on both expectancy and SCR measures. Thus, when an AB compound was followed by shock (AB+), responding to B was attenuated by reinforced presentations of A (A+) outside the compound. This outcome occurred whether the A+ presentations were experienced before (forward blocking) or after (backward blocking) the compound trials. It should be noted that the presentation of A+ on the first test trial in Experiment 1, following AB+ trials, means that the design included a backward blocking component. However, although backward blocking may have contributed to the effect found in Experiment 1, it is highly unlikely that backward but not forward blocking was observed in the present studies. More important, these cue competition effects were seen only in participants who were instructed that outcomes would be additive (the additive group). Participants given non-additivity instructions (the non-additive group) showed neither forward nor backward blocking on either the expectancy or the SCR measure. It seems highly likely, therefore, that past failures to demonstrate blocking in human Pavlovian conditioning (Davey & Singh, 1988; Lovibond et al., 1988) were due to participants' assumptions of outcome non-additivity.

A model of Pavlovian conditioning based on propositional/inferential processes is consistent with the present effects. Such a model predicts blocking to be strongest when participants assume additivity. In the A+/AB+ design, this prediction follows because it is only logical to conclude that B is non-causal as, had it been causal, B would have been expected to increase the level of the outcome in some way when combined with A. Blocking in the additive group could be considered to be an example of the conditional inference of modus tollens. By modus tollens, given the premises "if p then q", and "not q", it can be inferred that "not p". In the additive group in the present experiments, participants might reason that, if both A and B were causal (p), then the AB compound would have been followed by a double shock (q). The compound was not followed by a double shock (not q), therefore, it is not the case that both A and B are causal (not p). As it is known that A is causal, it can be inferred that B is not causal. Participants in the non-additive group cannot make this inference, as the compound AB might be expected to be followed by a single shock whether B was causal or not. Finally, an inferential model, assuming that all information were readily available on which to base the inference, would make no distinction between a forward and backward blocking procedure in predicting reduced responding to B on test in the additive group.

In order to account for the present data, any competing model would have to be able to explain why blocking is most effective when outcome additivity is assumed, and be consistent with the finding of backward blocking in Experiment 2. It would appear that both the associative theories, such as that proposed by Rescorla and Wagner (1972), and the statistical theories

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

of Cheng (Cheng, 1997; Cheng & Holyoak, 1995) would need modification in order to account for the present data. These two models are considered in turn.

Associative theory

For many associative learning models that rely on a trial-by-trial learning mechanism, such as the Rescorla–Wagner model (Rescorla & Wagner, 1972), trial order is critical, and the absence of backward blocking is strongly predicted. For instance, according to the Rescorla–Wagner model, forward blocking occurs because, following A+ pretraining, the US is fully expected on AB+ trials (A predicts the US), and therefore little learning occurs on these latter trials. Backward blocking requires an extension of the traditional associative theory, such as that suggested by Dickinson and Burke (1996). Dickinson and Burke postulate that an inhibitory link will form on A+ trials between the associatively activated B node and the directly activated US node. This will reduce responding to B on test, the backward blocking effect. This kind of modification has been used to account for the backward blocking effects found in causality judgement tasks (Dickinson & Burke, 1996) and may equally apply to the present Pavlovian conditioning data.

The additivity effects present a greater problem. There is nothing in the associative theories to explain the failure of the general learning mechanism when non-additivity instructions are given prior to training. However, it might be argued that associative theory was only ever designed to apply to situations in which outcomes were additive. It is this additivity assumption that allows the model to explain effects such as overexpectation and summation. It is explicitly stated in the Rescorla–Wagner model that the associative strength that has accrued to A (V_A) and B (V_B) will be combined when A and B are presented together; $V_{AB} = V_A + V_B$. If A and B are both paired with the US (A+ and B+), the expected outcome of the compound AB must, if the model is correct, be greater (in magnitude or probability) than the outcome of both of the elements A or B (AB++). If cues are limited to an associative strength of λ, the maximum supportable by a single US, then the additivity assumption of the model is violated, and so the present data might be considered outside of its scope. However, this argument raises further questions. For instance, how do the verbal instructions given before training disable the associative mechanism, and, if the associative mechanism does not apply, how one might model the behaviour of participants given non-additivity instructions? It is not immediately clear how such questions might be answered.

Normative statistical models

Another way in which blocking might be affected by outcome additivity has been suggested by Cheng (1997), in her statistical model of human causal reasoning, the power PC model. The power PC model is based on the earlier Probabilistic Contrast Model (PCM) of Cheng and Holyoak (1995). In the PCM, the probability of the outcome in the presence of the cue is calculated and compared to the probability of the outcome in the absence of the cue. The resulting contingency metric is taken to reflect the degree to which the participants perceive the target cue to be causal. The more recent power PC model was developed as the result of a recognition that, in a blocking design, if both Cue A and the compound AB lead to the outcome with a probability of 1, then the causal status of B will be ambiguous (Cheng, 1997). To account for this, the model calculates a measure of "causal power". Causal power is the contingency metric

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

of the PCM, modified by the base rate of the effect. Most importantly for the present purposes, in the special case where the base rate probability of the outcome is 1 (thus outcome probability is necessarily non-additive) the causal power of the B cue remains undefined in the power PC model. Failure to define the causal power of B under these circumstances is meant to reflect the participant's perception that B is causally ambiguous.

The power PC model correctly predicts both forward and backward blocking. In addition, the power PC model explicitly predicts some failures to show a blocking effect in situations where outcomes are not additive with respect to probability. However, the power PC model does not allow that the outcome may vary along other dimensions than probability, such as magnitude. The blocking effects demonstrated in the present experiments are inconsistent with the power PC model because the shock occurred with a probability of 1 following the A element and the AB compound. According to the power PC model, the B element would be expected to be ambiguous following such training regardless of any manipulation of magnitude additivity.

One core feature of the power PC model is the manner in which causal strengths are calculated—the comparison of the probability of the outcome in the presence versus the absence of the target cue. Such a comparison would appear to be a candidate component of the propositional account of causal reasoning proposed here. However, it is clear that the mathematical model specified by Cheng (1997) falls short of a complete account of blocking. The most natural modification for the power PC model would appear to be to make the maximum possible level of the outcome, on each of the dimensions along which the outcome might vary, an input into the equation. The model would, of course, have to allow that such inputs could be derived from both direct experience and instruction.

In summary, the power PC model and Rescorla–Wagner learning models make their own assumptions about additivity. The power PC model assumes non-additivity of outcomes along all dimensions except probability, whereas the Rescorla–Wagner model assumes additivity along all dimensions. The present data suggest that the models cannot accommodate data from studies in which the participants engaged in the tasks do not share these assumptions.

Although the propositional reasoning account is more consistent with the present data than are the associative or the probabilistic accounts presented earlier, there are questions that remain to be addressed. First, what is the nature of this reasoning process? Both the associative and probabilistic accounts of causal reasoning provide rules by which precise cause–effect relations can be calculated. In addition, they make quantitative predictions as to the strength of a response under varying environmental contingencies. Existing theories suggest that propositional reasoning might be based on logical rules (e.g., Braine, 1978; Rips, 1983), or possibly the creation of mental models (Johnson-Laird, 1983). Such processes, supplemented by a mechanism by which the participant's confidence in a particular inference could be derived, might gain the level of predictive precision achieved by the probabilistic and associative accounts.

Second, how can a deductive process give rise to conditioned responding? An example of the way in which propositional knowledge may give rise to CRs has been suggested by Lovibond (in press). In his account, CRs are the result of an expectancy of the outcome. Thus, knowledge of the causal status of the world in propositional form interacts with the present state of affairs to produce an expectancy of a particular outcome, which in turn generates

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

anticipatory responses (US-appropriate CRs). That is to say, if (1) the participant believes that blue leads to shock, and (2) blue appears, then the participant will expect shock, and his/her SCR will rise as a consequence. Of course, a similar mechanism would also be required for any modified version of the probability-based models (Cheng, 1997; Cheng & Holyoak, 1995) to explain CR generation.

Finally, what implications would follow from a propositional model for the study of animal learning and its relation to human learning? Backward blocking is very rare in the animal learning literature, and non-human animals cannot be verbally instructed to change their assumptions of outcome additivity. Thus, there are clear differences between human and non-human animal learning. These differences might be viewed as qualitative or quantitative. Human learning might be thought to involve the manipulation of propositional information, qualitatively distinct from the trial-by-trial error-correction process underlying animal learning, such as that proposed by Rescorla and Wagner (1972). Many comparative psychologists may be uncomfortable with such a stark division. Human and animal cognition might be brought closer together by postulating a levels of learning hypothesis (e.g., Squire, 1994). Thus, humans and animals may share a low-level associative system, whereas humans uniquely posses a high-level propositional system. However, the present data do not support such a view; the postulated low-level associative system did not affect either shock expectancy or SCRs. Also, there is little other supporting evidence for the levels of learning hypothesis in its present form (Lovibond & Shanks, 2002). A third, and perhaps more extravagant, alternative is that animal learning results from reasoning processes similar to, but simpler than, those used by humans—that is, that the difference between humans and animals is quantitative rather than qualitative. In fact, it might be argued that the comparator model of Miller and colleagues, and especially its recent extension (Denniston, Savastano, & Miller, 2001), represents a model of very simple reasoning. Although the comparator model is thought to describe an associative process, it could also be viewed as a conditional inference mechanism.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

REFERENCES

Braine, M.D.S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, *85*, 1–21.

Chapman, G.B., & Robbins, S.J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, *18*, 537–545.

Cheng, P.W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.

Cheng, P.W., & Holyoak, K.J. (1995). Complex adaptive systems as intuitive statisticians: Causality, contingency and prediction. In J.A. Meyer & H. Roitblat (Eds.), *Comparative approaches to cognition* (pp. 271–302). Cambridge, MA: MIT Press.

Davey, G.C.L., & Singh, J. (1988). The Kamin "blocking" effect and electrodermal conditioning in humans. *Journal of Psychophysiology*, *2*, 17–25.

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *Quarterly Journal of Experimental Psychology*, *55A*, 965–985.

Denniston, J.C., Savastano, H.I., & Miller, R.R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R.R. Mowrer & S.B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65–117). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, *49B*, 60–80.

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

Hinchy, J., Lovibond, P.F., & Ter-Horst, K.M. (1995). Blocking in human electrodermal conditioning. *Quarterly Journal of Experimental Psychology*, *48B*, 2–12.

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Kamin, L.J. (1969). Predictability, surprise, attention and conditioning. In B.A. Campbell & R.M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kimmel, H.D., & Bevill, M.J. (1991). Blocking and unconditional response diminution in human classical autonomic conditioning. *Integrative Physiological and Behavioral Science*, *26*, 132–138.

Kimmel, H.D., & Bevill, M.J. (1996). Blocking and unconditional response diminution in human classical autonomic conditioning. *Integrative Physiological and Behavioral Science*, *31*, 18–43.

Lovibond, P.F. (1992). Tonic and phasic electrodermal measures of human aversive conditioning with long duration stimuli. *Psychophysiology*, *29*, 621–632.

Lovibond, P.F. (in press). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Lovibond, P.F., Been, S., Mitchell, C.J., Bouton, M.E., & Frohardt, R. (2001). *Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude*. Manuscript submitted for publication.

Lovibond, P.F., & Shanks, D.R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 3–26.

Lovibond, P.L., Siddle, D.A.T., & Bond, N. (1988). Insensitivity to stimulus validity in human Pavlovian conditioning. *Quarterly Journal of Experimental Psychology*, *40B*, 377–410.

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

Miller, R.R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, *125*, 370–386.

Miller, R.R., & Schachtman, T.R. (1985). The several roles of context at the time of retrieval. In P.D. Balsam & A. Tomie (Eds.), *Context and learning* (pp. 167–194). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

O'Brien, R.G., & Kaiser, M.K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, *97*, 316–333.

Pearce, J.M., & Hall, G. (1980). A model of Pavlovian learning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*, *87*, 532–552.

Pellon, R., & Montano, J.M.G. (1990). Conditioned stimuli as determinants of blocking in human electrodermal conditioning. In P.J. Drenth, J.A. Sergeant, & R.J. Takes (Eds.), *European perspectives in psychology Vol. 2* (pp. 409–423). Chichester, UK: John Wiley & Sons.

Pellon, R., Montano, J.M.G., & Sanchez, P. (1995). Blocking and electrodermal conditioning in humans. *Psicologica*, *16*, 321–329.

Price, P.C., & Yates, J.F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition*, *21*, 561–572.

Razran, G. (1955). Conditioning and perception. *Psychological Review*, *62*, 83–95.

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* (pp. 64–99). New York: Appleton-Century-Crofts.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, *90*, 38–71.

Shanks, D.R. (1985). Forward and backward blocking in human contingency judgments. *Quarterly Journal of Experimental Psychology*, *37B*, 1–21.

Squire, L.R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D.L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 203–231). Cambridge, MA: MIT Press.

Sutherland, N.S., & Mackintosh, N.J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015

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: Conditioned inhibition* (pp. 223–266). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wasserman, E.A., & Berglan, L.R. (1998). Backward blocking and recovery from overshadowing in human causal judgment: The role of within compound associations. *Quarterly Journal of Experimental Psychology*, *51B*, 121–138.

Original manuscript received 2 January 2002
Accepted revision received 20 March 2002

Downloaded by [The UC Irvine Libraries] at 16:30 16 December 2015