University of Groningen

Learned predictiveness influences automatic evaluations in human contingency learning

Le Pelley, M. E.; Calvini, G.; Spears, R.

Published in:
Quarterly Journal of Experimental Psychology

DOI:
[10.1080/17470218.2012.708760](https://doi.org/10.1080/17470218.2012.708760)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Le Pelley, M. E., Calvini, G., & Spears, R. (2013). Learned predictiveness influences automatic evaluations in human contingency learning. *Quarterly Journal of Experimental Psychology*, 66(2), 217-228.
<https://doi.org/10.1080/17470218.2012.708760>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Learned predictiveness influences automatic evaluations in human contingency learning

M. E. Le Pelley^{1,2}, G. Calvini³, and R. Spears⁴

¹School of Psychology, Cardiff University, Cardiff, UK

²School of Psychology, University of New South Wales, Sydney, NSW, Australia

³School of Psychology, Plymouth University, Plymouth, UK

⁴Department of Psychology, University of Groningen, Groningen, The Netherlands

Experiments used an affective priming procedure to investigate whether evaluative conditioning in humans is subject to bias as a consequence of differences in the learned predictiveness of the cues involved. Experiment 1, using brief prime presentation, demonstrated stronger affective priming for cues that had been predictive of a neutral attribute prior to evaluative conditioning than for those that had been nonpredictive. Experiment 2, using longer prime presentation, found a reversed priming effect for previously predictive cues but not for previously nonpredictive cues. The implication is that the effect observed with brief prime presentation reflects the operation of fast-acting, automatic evaluation mechanisms and hence that evaluative conditioning can be biased by our previous learning about the predictiveness of cues.

Keywords: Evaluative conditioning; Associative learning; Human contingency learning; Bias; Attention.

Empirical evidence suggests that the majority of our likes and dislikes are learned, rather than innate (e.g., Rozin & Millman, 1987), but there is currently no general consensus regarding the mechanisms by which this learning occurs. This issue has motivated the study of evaluative conditioning: “changes in the liking of a stimulus that result from pairing that stimulus with other positive or negative stimuli” (De Houwer, Thomas, & Baeyens, 2001, p. 853). In other words, evaluative conditioning describes an associative transfer of valence from one stimulus to

another as a result of experience of pairings of those stimuli.

Notably, it has been suggested that the processes underlying evaluative conditioning might be different to those recruited by nonevaluative learning—that is, learning of a predictive relationship between stimuli that does not involve a change in liking (Baeyens & De Houwer, 1995; De Houwer et al., 2001; Klauer & Meiser, 2000; Martin & Levey, 1994). This suggestion has been fuelled by the observation of certain functional differences between the factors influencing

Correspondence should be addressed to M. E. Le Pelley, School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: m.lepelley@unsw.edu.au

This work was supported by Economic and Social Research Council Grant RES-062-23-0182.

evaluative and nonevaluative conditioning. For example, it has been claimed that, unlike nonevaluative learning, evaluative conditioning is not reduced by extinction (Diaz, Ruiz, & Baeyens, 2005; Dwyer, Jarratt, & Dick, 2007; but see Lipp, Oughton, & LeLievre, 2003), is not sensitive to the degree of contingency between the conditioned and unconditioned stimulus (Baeyens, Hermans, & Eelen, 1993), and is not susceptible to cue competition effects (Dwyer et al., 2007; but see Lipp, Neumann, & Mason, 2001; Purkis & Lipp, 2010). This article probes further the functional characteristics of evaluative conditioning by examining the extent to which it is prone to modulation by the *learned predictiveness* of the stimuli involved.

It is now well established that prior learning about the predictiveness of a stimulus influences the rate of future learning about that stimulus, both in humans and in nonhuman animals (see Le Pelley, 2004; Mitchell & Le Pelley, 2010, for reviews). The typical finding in humans is that participants learn more rapidly about stimuli (or cues) that have previously been experienced as predictive of events of significance than about cues previously experienced as nonpredictive (e.g., Bonardi, Graham, Hall, & Mitchell, 2005; Kruschke, 1996; Le Pelley & McLaren, 2003; Le Pelley, Reimers, et al., 2010; Le Pelley, Turnbull, Reimers, & Knipe, 2010; Owen, Roberts, Polkey, Sahakian, & Robbins, 1991). In all of these experiments, differences in the previously experienced predictiveness of cues produced biases in what was later learnt about those cues. According to the attentional interpretation of these findings (Kruschke, 2003; Mackintosh, 1975), people learn to pay more attention to cues that are predictive of important events and to ignore cues that predict nothing of importance. Consequently, when these cues are presented later, participants will learn more rapidly about the previously predictive cues because they are now paying more attention to them.

Notably, previous studies of learned predictiveness in humans have almost exclusively used nonevaluative learning procedures; studies have assessed the learning of the predictive relationships

between cues and outcomes, rather than looking at whether learned predictiveness can bias the formation of evaluations of cues. One exception is a recent study by Le Pelley, Reimers, et al. (2010); Table 1 shows their design. On each trial of Stages 1 and 2, participants read a description of an individual who belonged to two gangs (e.g., the Thorps, Quents, etc., labelled G1–G8 in Table 1). During Stage 1, following this description, participants had to decide which of two pictures, differing only in clothing colour (green or yellow), showed this individual. For individuals belonging to Gang G1 or G4, the correct figure was always in green; for individuals belonging to G2 or G3, the correct figure was in yellow. Membership of Gangs G5–G8, in contrast, provided no basis for discrimination. Half of the individuals belonging to G5 wore green, while the other half wore yellow; the same applied to Gangs G6–G8. Each combination in Table 1 was shown several times, allowing participants to learn that Gangs G1–G4 predicted colour while G5–G8 were nonpredictive.

In Stage 2, following the description of each individual, participants read a statement describing a positive behaviour (e.g., “helped friends to move house”) or negative behaviour (e.g., “trespassed on private property”) performed by that individual.

Table 1. *Experimental design*

<i>Stage 1</i>	<i>Stage 2</i>
G1, G5 → green	G1, G7 → positive
G1, G6 → green	G2, G8 → negative
G2, G5 → yellow	G3, G5 → positive
G2, G6 → yellow	G4, G6 → negative
G3, G7 → yellow	
G3, G8 → yellow	
G4, G7 → green	
G4, G8 → green	

Note: Symbols G1–G8 represent different gangs to which target individuals were described as belonging. “Green” and “yellow” refer to colours of clothing worn by these target individuals. For example, “G1, G5 → green” indicates that the correct figure that participants should choose for a target individual described as belonging to Gangs G1 and G5 is the figure in green. “Positive” and “negative” refer to the affective valence of behaviour statements that were attributed to target individuals in Stage 2.

Each gang combination shown in Table 1 was again experienced several times. Individuals defined by combinations G1,G7 and G3,G5 performed positive behaviours; individuals defined by G2,G8 and G4,G6 performed negative behaviours. Thus gangs paired with green in Stage 1 were equally likely to perform positive or negative behaviours in Stage 2; the same applied for gangs paired with yellow in Stage 1. This rendered behaviour valence statistically independent of clothing colour. For example, knowing that members of a particular gang wore green told a perceiver nothing about the valence of behaviours performed by members of that gang.

The Stage 2 behaviour statements should have led participants to develop evaluations of the gangs. The question of interest was whether the strength of these evaluations was biased by participants' previous experience of the predictiveness of these gangs with respect to clothing colour. On the basis of previous studies of learned predictiveness in human contingency learning, we would expect more rapid learning (and hence formation of stronger evaluations) regarding gangs experienced as predictive in Stage 1 (G1–G4), than non-predictive gangs (G5–G8).

Le Pelley, Reimers, et al. (2010) assessed evaluations by simply having participants rate how much they liked members of each gang, on a scale from 0 to 10. As anticipated, likeability ratings were more extreme (suggesting stronger evaluations) for previously predictive gangs than for previously non-predictive gangs. That is, previous learning about the predictiveness of the gangs produced a systematic bias in Stage 2 learning.

Le Pelley, Reimers, et al.'s (2010) data are clearly consistent with the suggestion that learned predictiveness modulates evaluative conditioning in the same way as it modulates nonevaluative learning and hence support the idea of a common functional dependency. One problem remains, however. The use of an explicit, rating-based test means that the observed patterns of judgements may not reflect participants' "real" evaluations of the gangs. For example, when cued for a likeability rating for the Thorps, a participant might explicitly recall that all members of the Thorps encountered during Stage 2 were paired with negative

behaviours. Consequently, they might reason that the experimenter expects them to give a low rating even though, in fact, they neither like nor dislike members of the Thorps. On this view, participants' ratings reflect a demand characteristic (resulting from reasoning based on nonevaluative, episodic memories of Stage 2), rather than being a true measure of evaluative conditioning (essentially a semantic memory effect, in which gang names take on positive or negative connotations).

Fortunately, there is a straightforward way to distinguish these accounts, by using an indirect test to measure participants' *automatically elicited* evaluations of the gangs. Here we use the affective priming task (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Many studies using this task have demonstrated that a valenced target stimulus is responded to more rapidly after presentation of an affectively congruent prime than after an incongruent prime (see Klauer & Musch, 2003, for a review). For example, participants are typically faster to respond that the target word *honest* has positive valence if it is preceded by the positively valenced prime *happiness* than if it is preceded by the negatively valenced prime *murder*. More recent studies have used affective priming to measure evaluative conditioning of initially neutral stimuli paired with positive or negative events (e.g., De Houwer, Hermans, & Eelen, 1998; Hermans, Spruyt, & Eelen, 2003; Olson & Fazio, 2002; Stahl, Unkelbach, & Corneille, 2009). Here we use affective priming to assess evaluative conditioning of the gang names resulting from Stage 2 pairings with valenced behaviour in the design in Table 1.

While the affective priming task is not completely out of participants' conscious control (see discussion following Experiment 1), it is clearly less susceptible to the influence of strategic memory processes than are explicit likeability ratings (Klauer & Teige-Mocigemba, 2007). Participants are required to respond to the valence of the target and not to the prime. Since there is no explicit demand to evaluate the prime, there is no call for participants to reason about its status based on conscious memories of the events with which it has been paired. Neither is there any advantage in

doing so; since primes are followed by incongruent targets as often as by congruent targets, explicitly evaluating the prime will not, on average, speed responses to the target. Finally, there is little opportunity for controlled processing of the prime; the stimulus onset asynchrony (SOA) between prime and target presentation is typically short (300 ms in Experiment 1). Following a gradualist approach to automaticity, the activations occurring under such restrictive time conditions are considered faster, more efficient, and less controllable (Moors & De Houwer, 2006).

More generally, then, use of an affective priming measure allows us to investigate the extent to which learning can produce changes in behaviour at an automatic level. Previous studies have used priming to assess the *existence* of evaluative conditioning (e.g., De Houwer et al., 1998; Hermans et al., 2003; Olson & Fazio, 2002; Stahl et al., 2009). In other words, these studies indicate that pairing a cue with a positive or negative event can produce a change in the automatic evaluation of that cue (although see De Houwer, 2006; Klauer & Teige-Mocigemba, 2007). The current experiments go beyond this prior research by using affective priming as a tool to investigate the factors that modulate the learning underlying these changes in automatic behaviour. This issue is taken up again in the General Discussion.

EXPERIMENT 1

Experiment 1 followed the training in Table 1 with an affective priming task, to assess the evaluative conditioning that occurred during Stage 2. If evaluative conditioning is subject to the biasing influence of learned predictiveness, there will be greater affective priming to gangs experienced as predictive of clothing colour during Stage 1 (G1–G4) than to nonpredictive gangs (G5–G8).

Method

Participants, apparatus and stimuli

Sixteen Cardiff University undergraduates (5 male) participated for course credit. The experiment

programme was written in Microsoft Visual Basic 6 and was run on standard PCs. The eight gang names—six-letter nonsense syllables (e.g., Thorps, Quents)—were randomly assigned to designations G1–G8 for each participant. Stage 1 options were pictures of a male differing only in clothing colour: one in green, the other in yellow. Sentences describing positive and negative behaviours used in Stage 2 were those of Le Pelley, Reimers, et al. (2010).

Procedure

On each Stage 1 trial, participants read a *person description* stating the two gangs to which a person belonged and then used the mouse to select which of the two pictures showed this person with immediate corrective feedback. Stage 1 comprised 16 blocks, featuring each of the eight trial types in Table 1 once per block in random order. The order of the two gangs in the person description was counterbalanced across blocks for each trial type—for example, for trial type G1, G5 → green, there were eight presentations with G1 before G5, and eight with G5 before G1 (the order of these presentations was randomized).

On each Stage 2 trial, participants read a person description, followed by a behaviour statement—for example, “He trespassed on private property”, before moving to the next trial. Stage 2 comprised 10 blocks, featuring each of the four trial types once per block in random order. Presentation order was counterbalanced as for Stage 1.

Affective priming task instructions stated that a gang name and a valenced word would be presented sequentially in the centre of the screen, and participants were to respond to the second word’s valence as rapidly as possible, using the keyboard. Each trial began with a central fixation cross, replaced after 700 ms by a 200-ms prime presentation. Primes and targets were separated by a 100-ms blank screen, giving 300-ms SOA. Assignment of response keys (C and M) to valence responses (positive/negative) was randomly determined for each participant. Correct trials were not followed by feedback; on incorrect trials, the message “Incorrect” appeared in the centre of the screen for 1,000 ms, and the computer beeped. If no response was registered within 3,000 ms, a timeout occurred; the

message “You took too long” appeared in the centre of the screen for 1,000 ms, and the computer beeped. The intertrial interval was 2,000 ms.

Participants first received 16 practice trials, during which each gang primed a positive target (*beautiful*) and a negative target (*miserable*) once. In the three experimental blocks that followed, each gang appeared eight times as a prime, once before each of four positive targets (*delightful*, *wonderful*, *appealing*, *terrific*), and once before each of four negative targets (*dreadful*, *frightful*, *disgusting*, *terrible*), in random order.

Results and discussion

There could be an influence of Stage 1 predictiveness on evaluative conditioning only if participants learned the Stage 1 gang–colour relationships. Following Le Pelley and McLaren (2003), a criterion of 60% correct responses averaged across all Stage 1 trials was imposed. Two participants failed this criterion and were excluded from further analysis. For remaining participants, mean percentage correct rose steadily during Stage 1, reaching 85.7% in Block 16.

Data from priming trials with responses faster than 200 ms (0.04% of trials) or slower than 1,500

ms (0.4%) were discarded to reduce the influence of outlier responses. After preliminary analysis revealed no significant effects on accuracy, errors were also excluded (5.7%). Gangs paired with positive behaviours in Stage 2 (G1/G3/G5/G7) were termed positive primes; gangs paired with negative behaviours (G2/G4/G6/G8) were termed negative primes. Trials could be affectively congruent (positive prime and positive target; negative prime and negative target), or incongruent (positive prime and negative target; negative prime and positive target). Figure 1A shows mean response latency for previously predictive gangs (G1–G4) and previously nonpredictive gangs (G5–G8). Analysis of variance (ANOVA) with factors of predictiveness and congruence revealed a main effect of congruence, $F(1, 13) = 6.30, p < .05$, with faster responses on congruent than on incongruent trials. This *positive compatibility effect* demonstrates affective priming and shows that Stage 2 training produced evaluative conditioning of the gangs. Priming was greater for previously predictive gangs than for previously nonpredictive gangs: Predictiveness $\times$ Congruence interaction, $F(1, 13) = 6.14, p < .05$. Simple effects analysis revealed a significant effect of congruence for previously predictive gangs, $F(1, 13) = 21.2$,

Figure 1. Response latencies on congruent and incongruent trials, averaged across the three blocks of the priming test in (A) Experiment 1, which had a short prime–target stimulus onset asynchrony (SOA) of 300 ms, and (B) Experiment 2, which had a longer SOA of 1,000 ms. Data are shown separately for gangs (which acted as primes) that had been predictive of clothing colour during Stage 1 (Gangs G1–G4) and those that had been nonpredictive of clothing colour (Gangs G5–G8).

$p < .001$, but not for previously nonpredictive gangs, $F < 1$. Thus evaluative conditioning during Stage 2 proceeded more rapidly for gangs that previously predicted clothing colour than for nonpredictive gangs.

There is a caveat, however. While affective priming is less susceptible to strategic demand processes than are likeability ratings, evidence from Klauer and Teige-Mocigemba (2007) indicates that priming is not entirely immune to strategic influences. These researchers showed that explicit instruction to participants regarding the likely valence of the target that would follow a particular prime influenced the affective priming observed to that prime, independent of the prime's own valence. For example, telling participants that negative targets were more likely to appear following a particular positive prime (e.g., *happiness*) eliminated the normal pattern of faster responses to positive targets following this positive prime. In the current Experiment 1, participants received no explicit information on the relationship between primes and targets, and each prime was followed equally often by positive and negative targets. Hence the current results cannot be explained by the type of strategic process identified by Klauer and Teige-Mocigemba. However, their work (see also De Houwer, 2006) raises the possibility that some other form of controlled process could explain the Experiment 1 data. Perhaps participants do strategically evaluate the prime despite having no incentive, and little time, to do so. Experiment 2 was therefore conducted to test this possibility.

EXPERIMENT 2

If Experiment 1's findings reflect conscious response strategies, then providing more time to process prime–target relations (longer SOA) should produce stronger or at least similar effects, assuming that such controlled strategies are time consuming (Fazio, 2001; Posner & Snyder, 1975).

Method

Participants, apparatus, stimuli, and procedure

Forty undergraduates participated. The only difference from Experiment 1 was that primes appeared

for 900 ms, followed by 100-ms blank screen, giving 1,000-ms SOA.

Results and discussion

Thirteen participants failed to achieve 60% correct across Stage 1 and were excluded from further analysis. Mean accuracy for remaining participants reached 95.4% in Block 16.

Data were discarded from priming trials with responses faster than 200 ms (0.2% of trials) and slower than 1,500 ms (1.3%). After analysis revealed no significant effects on accuracy, errors were excluded (4.5%). Figure 1B shows mean response latency for the remaining trials. ANOVA with factors of predictiveness and congruence revealed a main effect of congruence, $F(1, 26) = 6.45$, $p < .05$, with faster responses on incongruent trials than on congruent trials. This *negative compatibility effect* is opposite to the pattern observed in Experiment 1. There was some evidence that this negative compatibility effect was greater for previously predictive gangs than for previously nonpredictive gangs. While the Predictiveness $\times$ Congruence interaction was nonsignificant, $F(1, 26) = 1.29$, $p = .27$, simple effects analysis revealed a significant negative compatibility effect for previously predictive gangs, $F(1, 26) = 7.84$, $p < .01$, but not for previously nonpredictive gangs, $F < 1$.

In summary, then, the results of Experiments 1 and 2 each showed significant patterns but in opposite directions. In order to more closely compare the findings of these studies we conducted a between-experiments comparison. These experiments were conducted by the same experimenter using the same apparatus, with identical instructions to participants, and recruited participants from the same undergraduate population. Note, however, that as with any between-experiments comparison, the findings of this analysis should be treated with some caution as participants were not randomly assigned to experiments (Experiment 1 was run prior to Experiment 2, with a one-day gap between the two studies). Indeed, it could be argued that the numerical difference in proportion of participants excluded for failing to reach the

criterion of 60% correct in Stage 1 (12.5% of participants in Experiment 1 and 32.5% in Experiment 2) indicated that the two participant samples may not be entirely equivalent, although this difference was not significant, $p = .19$ by Fisher's exact test.

With the appropriate caution in mind, ANOVA using data for accuracy during Stage 1 training, with factors of experiment and block, found a main effect of block, $F(15, 585) = 20.0$, $p < .001$, with accuracy increasing across blocks. However, neither the main effect of experiment nor the Experiment $\times$ Block interaction was significant, higher $F = 1.19$, $p = .27$. This is not surprising given that Stage 1 training was identical in Experiments 1 and 2, but the finding that participants performed similarly increases confidence in the comparability of the two studies.

Of more interest, ANOVA using the affective priming data with factors of experiment, predictiveness, and congruence found a significant Experiment $\times$ Congruence interaction, $F(1, 39) = 12.98$, $p < .001$, reflecting the fact that Experiment 1 demonstrated a positive compatibility effect (faster responses on congruent than incongruent trials), while Experiment 2 found a negative compatibility effect. The three-way interaction was also significant, $F(1, 39) = 4.31$, $p < .05$, since there was a difference in compatibility effects in the two experiments for previously predictive gangs, but not for previously nonpredictive gangs (no priming occurred in either experiment for previously nonpredictive gangs). The main effect of experiment approached significance, $F(1, 39) = 3.79$, $p = .059$, with a trend towards slower responses in Experiment 2.

GENERAL DISCUSSION

While Experiment 1 demonstrated a positive compatibility effect, Experiment 2 found a negative compatibility effect. Making it easier for participants to use strategic processes, by lengthening SOA, did not result in stronger or even similar affective priming; instead it reversed the effect. The negative compatibility effect obtained in

Experiment 2 suggests that whatever processes are operating at long SOA in the priming task are not simply "more of the same" processes that are driving responses under short SOA. More specifically, this dissociation across studies indicates that the affective priming observed with short SOA in Experiment 1 reflects the operation of a fast-acting, automatic evaluation process, rather than being the product of conscious response strategies. That is, the affective priming in Experiment 1 demonstrates genuinely evaluative conditioning of the cues, as opposed to reasoning based on nonevaluative, episodic memories of the information experienced during Stage 2.

Experiment 2, using a longer SOA than Experiment 1, found a negative compatibility effect that was significant for previously predictive gangs but not for previously nonpredictive gangs. We are not the first to report data showing a positive compatibility effect at short SOA and a negative compatibility effect at long SOA in affective priming (Klauer, Roßnagel, & Musch, 1997), or in the wider motor priming literature (see Boy & Sumner, 2010; Eimer & Schlaghecken, 2003; Sumner, 2007). However, we are the first to demonstrate this pattern in the context of evaluative conditioning; previous studies have used stimuli that are semantically related (e.g., *happiness* as a prime for *honest*, or left-pointing arrows as a prime for a left-hand keypress). This suggests that the priming observed for "episodically related" stimuli (as used here) is qualitatively similar to that for semantically related stimuli. Briefly, it is thought that a negative compatibility effect occurs when there is time for the partial activation caused by the prime to be detected and suppressed by a self-inhibition mechanism; effectively, participants overcorrect for the influence of the prime when SOA is long (Boy & Sumner, 2010; Eimer & Schlaghecken, 2003; Sumner, 2007). The trend towards slower responses in Experiment 2 than Experiment 1 tallies with this suggestion, with the difference reflecting the extra time needed to overcome response suppression. The observation of a significant negative compatibility effect only for previously predictive gangs also fits with this account: Since previously nonpredictive

gangs do not produce a positive compatibility effect at short SOA, the small response activation that they generate will not produce significant suppression at long SOA, and so no negative compatibility effect will occur. Whether suppression results from strategic processes operating during the long SOA (Klauer et al., 1997), or another automatic mechanism (Boy & Sumner, 2010; Eimer & Schlaghecken, 2003), remains a matter of debate. Either way, as noted earlier, the finding that lengthening SOA reverses the priming effect indicates that the biasing influence of previously learned predictiveness in Experiment 1 cannot result wholly from controlled reasoning.

Previous studies of the learned predictiveness effect in evaluative conditioning, reported by Le Pelley, Reimers, et al. (2010), used an explicit measure wherein participants rated the likeability of the different gangs. In all of these studies, participants' ratings were congruent with their experience of each gang's valence during Stage 2—that is, gangs paired with positive behaviours in Stage 2 generally elicited higher likeability ratings than those paired with negative behaviours. In the terminology of the current article, these studies demonstrated a positive compatibility effect. There was no time pressure for participants to provide their ratings, such that controlled processes could presumably contribute to their behaviour. But we have argued above that the influence of controlled processes in the current Experiment 2 led to a *negative* compatibility effect. So why did these previous studies using likeability ratings detect a positive compatibility effect?

In the affective priming task used in the current experiments, participants were required to respond to the target word and *not* to the prime (the gang name). Consequently in this priming task the role of controlled processes will be to suppress any automatic activation caused by presentation of the gang name, since responding on the basis of this activation is inappropriate. This will produce a negative compatibility effect for the reasons outlined earlier. In contrast, in the explicit rating task used by Le Pelley, Reimers, et al. (2010), participants were required to respond to the valence of the gang name itself. There was therefore no reason for

them to use controlled processes to suppress any automatic activation generated by presentation of the gang name, since responding on the basis of this automatic activation was entirely appropriate in this task. Indeed, if anything, controlled processes akin to demand characteristics might be expected to enhance responding consistent with this automatic activation. Hence the format of the task used at test—whether participants are required to respond directly according to the valence of the gang name (as in a likeability rating task), or to withhold such responding (as in an affective priming task)—will be the crucial determinant of whether controlled processes act in concert with automatic processes to produce a positive compatibility effect, or act against them to produce a negative compatibility effect.

Learned predictiveness and attention

Experiment 1 found stronger evaluative conditioning for gangs that were previously predictive of a neutral attribute (clothing colour) than for those that were nonpredictive, even though clothing colour was statistically independent of, and hence unrelated to, behaviour valence. Experiment 2 also found a significant reversed priming effect only for previously predictive gangs, suggesting that these gangs were the only ones that generated sufficient automatic response activation to require suppression by controlled processes at long SOA.

The findings of both experiments therefore indicate that the extent of evaluative conditioning is subject to bias as a result of differences in the learned predictiveness of the cues involved, just as previous studies have demonstrated is the case for nonevaluative learning (e.g., Bonardi et al., 2005; Griffiths & Le Pelley, 2009; Kruschke, 1996; Le Pelley & McLaren, 2003; Le Pelley, Reimers, et al., 2010; Le Pelley, Turnbull, et al., 2010; Owen et al., 1991). This pattern of more rapid learning about previously predictive gangs is consistent with an interaction between attention and associative learning along the lines suggested by Mackintosh (1975). On this account, participants pay more attention to cues that they have learnt to be predictive during Stage 1, and this greater

attention in turn fuels more rapid evaluative conditioning with regard to these gangs during Stage 2 than to gangs experienced as nonpredictive in Stage 1. Other accounts are possible—for example, in terms of differences in memory representations (see Le Pelley, Reimers, et al., 2010, pp. 156–157). However, evidence from studies using a range of outcome measures, including eye gaze, electrophysiology, and susceptibility to the “attentional blink”, favours an interpretation of these findings in terms of an attentional bias (see Le Pelley, 2010).

The suggestion that the extent of evaluative conditioning to a cue is related to the amount of attention paid to that cue is consistent with the findings of previous studies by Field and Moore (2005), and Pleyers, Corneille, Yzerbyt, and Luminet (2009). These researchers demonstrated that drawing participants’ attention away from the cues of an evaluative conditioning procedure with an attention-demanding secondary task—counting down from 300 in intervals of 3 (Field & Moore), or a “two-back” memory task (Pleyers et al.)—led to a reduction in evaluative conditioning as measured by explicit likeability ratings. If one accepts the attentional interpretation of the current data (and those of Le Pelley, Reimers, et al., 2010), then our findings extend this previous work to a case in which the manipulation of attention to the cues involved in the evaluative conditioning task does not result from an explicit instruction to participants to engage in a distracting secondary task. Instead the difference in attention to these cues is learned by the participants themselves in a previous task that is (objectively) unrelated to the evaluative conditioning procedure, and persists to influence the specific pattern of evaluative conditioning within trials of this latter task. Moreover, the current results demonstrate that it is the learned predictiveness of the cues involved in the evaluative conditioning task that is the crucial determinant of the attention paid to them. And finally, the current findings demonstrate the influence of attention on evaluative conditioning using a more automatic measure (affective priming), which is less susceptible to the influence of demand characteristics than the explicit likeability ratings used in prior research.

Evaluative and nonevaluative learning

Given that a similar influence of learned predictiveness has been observed in several studies of nonevaluative learning in humans (e.g., Bonardi et al., 2005; Kruschke, 1996; Le Pelley & McLaren, 2003; Le Pelley, Reimers, et al., 2010; Le Pelley, Turnbull, et al., 2010), the implication is that, in this respect at least, evaluative and nonevaluative learning show a similar functional dependence.

More generally, the current findings add to the debate over whether evaluative learning and nonevaluative Pavlovian associative learning are underpinned by distinct learning processes (Davey, 1994; De Houwer, Baeyens, & Field, 2005; De Houwer et al., 2001; Dwyer et al., 2007; Lipp & Purkis, 2005). Notably, our findings provide a demonstration of *selective learning* in evaluative conditioning. That is, the learning occurring on each Stage 2 trial is not equal for both of the cues that are presented; instead more learning accrues to one cue (the one that was previously predictive in Stage 1) than the other. This is noteworthy, because it has previously been argued that a lack of susceptibility to selective learning effects is a functional difference that sets some forms of evaluative learning apart from nonevaluative learning. Using a procedure in which pictures of foods served as cues, and pictures of body shapes served as outcomes, Dwyer et al. (2007) found no evidence of a difference in the magnitude of evaluative conditioning between cues that were trained alone or in compound with another cue; that is, no overshadowing effect was observed. They concluded that their results were consistent with a mechanism for evaluative conditioning that was sensitive simply to the co-occurrence of the cue and outcome, and consequently not susceptible to selective learning. We note that Dwyer et al. did not test whether their procedure would yield overshadowing of nonevaluative learning, so it remains possible that their procedure was simply not sensitive or conducive to detecting overshadowing in general and hence may not be informative with regard to a functional difference between evaluative and nonevaluative learning. Notwithstanding this criticism, the current data and those of Le Pelley, Reimers,

et al. (2010; see also Purkis & Lipp, 2010) demonstrate clearly that selective learning effects *can* occur in the context of evaluative learning.

Prejudice and bias

In the language of social cognition, the current research extends previous work by Le Pelley, Reimers, et al. (2010) by demonstrating a systematic bias—a contingency illusion—in the formation of prejudice (i.e., involving some affective commitment) regarding fictitious social groups and not merely knowledge of group evaluation. Experiment 1 demonstrates that automatically elicited prejudice differs between previously predictive and previously nonpredictive gangs despite these gangs being paired with exactly the same valenced behaviours, such that the objective statistical gang–valence relationship is the same for both. And if the prejudices that we learn are systematically biased away from reality, then our future perception of groups and appraisal of their members will inevitably also be inaccurate.

The existence of biases should not be taken as an indictment of the learning system, however. While the design used here carefully arranges for the gang–event relationships in Stages 1 and 2 to be statistically independent, in the real world it is likely that previously predictive cues will also be accurate predictors of other information in future. Under such circumstances it is advantageous for information processing to be focused on these cues at the expense of others, as this will speed the learning of potentially important predictive information. An analogy can be drawn between the contingency illusion demonstrated here and the classic perceptual illusions implicating top-down processing, such as misapplied size constancy, illusory contours, or phoneme restoration (see Sekuler & Blake, 2005). The biasing influence of prior predictiveness (or, in our analogy, the influence of top-down processes in perception) might constitute a focusing and optimizing mechanism that distils what is learned (or perceived) down to the system’s “best guess” of what will be useful in future, but at the expense of meaning that what is

learned (or perceived) is no longer an objective reflection of reality.

Original manuscript received 9 December 2011

Accepted revision received 15 May 2012

First published online 4 September 2012

REFERENCES

- Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). *Behaviour Research and Therapy*, 33, 825–831.
- Baeyens, F., Hermans, D., & Eelen, P. (1993). The role of CS–US contingency in human evaluative conditioning. *Behaviour Research and Therapy*, 31, 731–737.
- Bonardi, C., Graham, S., Hall, G., & Mitchell, C. J. (2005). Acquired distinctiveness and equivalence in human discrimination learning: Evidence for an attentional process. *Psychonomic Bulletin & Review*, 12, 88–92.
- Boy, F., & Sumner, P. (2010). Tight coupling between positive and reversed priming in the masked prime paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 892–905.
- Davey, G. C. L. (1994). Is evaluative conditioning a qualitatively distinct form of classical conditioning? *Behaviour Research and Therapy*, 32, 291–299.
- De Houwer, J. (2006). Using the implicit association test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation*, 37, 176–187.
- De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion*, 19, 161–174.
- De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition and Emotion*, 12, 145–169.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853–869.
- Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion*, 19, 245–268.

- Dwyer, D. M., Jarratt, F., & Dick, K. (2007). Evaluative conditioning with foods as CSs and body shapes as USs: No evidence for sex differences, extinction, or overshadowing. *Cognition and Emotion*, 21, 281–299.
- Eimer, M., & Schlaghecken, F. (2003). Response facilitation and inhibition in subliminal priming. *Biological Psychology*, 64, 7–26.
- Fazio, R. H. (2001). On the automatic activation of attitudes. *Cognition & Emotion*, 15, 115–141.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.
- Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion*, 19, 217–243.
- Griffiths, O., & Le Pelley, M. E. (2009). Attentional changes in blocking are not a consequence of lateral inhibition. *Learning & Behavior*, 37, 27–41.
- Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion*, 17, 83–99.
- Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin*, 26, 1074–1093.
- Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ: Lawrence Erlbaum Associates.
- Klauer, K. C., Roßnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23, 246–255.
- Klauer, K. C., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology*, 43, 648–655.
- Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, 8, 225–247.
- Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, 12, 171–175.
- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*, 57B, 193–243.
- Le Pelley, M. E. (2010). Attention and human associative learning. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 187–215). Oxford, UK: Oxford University Press.
- Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *Quarterly Journal of Experimental Psychology*, 56B, 68–79.
- Le Pelley, M. E., Reimers, S. J., Calvini, G., Spears, R., Beesley, T., & Murphy, R. A. (2010). Stereotype formation: Biased by association. *Journal of Experimental Psychology: General*, 139, 138–161.
- Le Pelley, M. E., Turnbull, M. N., Reimers, S. J., & Knipe, R. L. (2010). Learned predictiveness effects following single-cue training in humans. *Learning & Behavior*, 38, 126–144.
- Lipp, O. V., Neumann, D. L., & Mason, V. (2001). Stimulus competition in affective and relational learning. *Learning and Motivation*, 32, 306–331.
- Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation*, 34, 219–239.
- Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition & Emotion*, 19, 269–282.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298.
- Martin, I., & Levey, A. (1994). The evaluative response: Primitive but necessary. *Behaviour Research and Therapy*, 32, 301–305.
- Mitchell, C. J., & Le Pelley, M. E. (Eds.). (2010). *Attention and associative learning: From brain to behaviour*. Oxford, UK: Oxford University Press.
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132, 297–326.
- Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition*, 20, 89–103.
- Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., & Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set shifting performance following frontal-lobe excisions, temporal-lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia*, 29, 993–1006.
- Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology: Animal Behavior Processes*, 35, 279–285.

- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solo (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Purkis, H. M., & Lipp, O. V. (2010). Stimulus competition in pre/post and online ratings in an evaluative learning design. *Learning and Motivation*, 41, 84–94.
- Rozin, P., & Millman, L. (1987). Family environment, not heredity, accounts for family resemblances in food preferences and attitudes: A twin study. *Appetite*, 8, 125–134.
- Sekuler, R., & Blake, R. (2005). *Perception*. New York, NY: McGraw-Hill.
- Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, 97, 404–420.
- Sumner, P. (2007). Negative and positive masked-priming: Implications for motor inhibition. *Advances in Cognitive Psychology*, 3, 317–326.