Cognitive Processes in Extinction

Peter F. Lovibond

Learn. Mem. 2004 11: 495-500
Access the most recent version at doi:10.1101/lm.79604

Review

Cognitive Processes in Extinction

Peter F. Lovibond

School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia

Human conditioning research shows that learning is closely related to consciously available contingency knowledge, requires attentional resources, and is influenced by language. This research suggests a cognitive model in which extinction consists of changes in contingency beliefs in long-term memory. Laboratory and clinical evidence on extinction is briefly reviewed, and it is concluded that the evidence supports the cognitive position. There is little evidence for a separate, noncognitive conditioning system. The primary implication for neural analysis is that learning and extinction are unlikely to be reducible to direct connections in which one stimulus simply activates or inhibits the memory representation of another. Rather, an adequate neural model will involve the integration of both low-level and high-level systems, including attention, representation of stimulus relations in long-term memory, and a dynamic performance mechanism based on anticipation, not just activation.

Extinction refers to the loss of an associatively based behavior when the associative relationships that generated the original learning have been changed. For example, in Pavlovian conditioning, the conditioned response (CR) declines over trials when the conditioned stimulus (CS) is no longer followed by the unconditioned stimulus (US). The way in which we view extinction necessarily follows from the way in which we conceptualize the initial conditioning. Psychologists, neuroscientists, and lay people alike share a common conceptualization of conditioning: It is seen as an automatic, unconscious, and low-level process that establishes excitatory or inhibitory associations between CS and US nodes in memory. In humans, conditioning is almost universally agreed to involve separate mechanisms from those involved in reasoning, language, and consciousness. Accordingly, extinction is also seen as an automatic and unconscious process that either reverses the original learned associations, or establishes new competing associations. But what if this traditional view of conditioning and extinction is wrong?

In fact, research on conditioning in humans suggests it to be an entirely different process from that suggested by the traditional view. Evidence for a separate conditioning mechanism that is independent of higher cognitive processes has been remarkably difficult to obtain (Brewer 1974; Dawson and Schell 1985; Lovibond and Shanks 2002). Instead, conditioning in humans appears to be closely tied to attention, consciousness, and language. In this article, I will briefly summarize this research, then focus specifically on extinction in human conditioning, and finally consider the implications for research into the neural basis of conditioning and extinction in both animals and humans.

Human Pavlovian Conditioning

As in animals, Pavlovian conditioning in humans involves pairing an initially neutral stimulus, such as a picture or a tone (the CS), with a stimulus of some significance (the US). Learning is indexed by the development of responses (CRs) to the CS that are the same as or similar to those elicited by the US. The most commonly used procedure is autonomic conditioning, in which the US is electric shock and the CR is a psychophysiological measure of autonomic arousal, usually skin conductance. Another popular procedure is eyeblink conditioning, in which the US is an airpuff to the eye and the CR is eyelid closure. In each of these procedures, it is common to use a differential conditioning design in which one CS (designated CS+) is paired with the US and a second CS (designated CS−) is paired with the absence of the US in order to control for nonassociative influences on responding.

Two types of self-report data may also be collected. First, participants may be asked in a postexperimental interview about their explicit knowledge of the relationships (contingencies) between the CSs and the US. Second, they may be asked to provide online ratings of US expectancy during presentation of the CS. The traditional view of conditioning assumes that self-reported knowledge derives from higher-order cognitive processes, whereas CRs are generated by a separate, low-level mechanism. Thus, the traditional view predicts that CRs will follow quite different principles from self-report measures (see Razran 1955; Squire 1994). However, reviews of this literature have consistently concluded that the two types of measure are closely related.

First, differential conditioning is only observed in participants who are aware of the relationship between the CSs and the US, as defined by their ability to verbalize those relationships (see Brewer 1974; Davey 1987; Lovibond and Shanks 2002). If conditioning was independent of the cognitive processes involved in conscious knowledge, it should be observed regardless of whether conscious learning has also occurred. Second, conditioning requires attentional resources. If attention is diverted by a secondary or masking task, not only is conscious learning slowed but so also is the development of CRs (see Dawson 1970). Third, ratings of US expectancy show a highly similar pattern to CRs on a trial by trial basis: They show similar acquisition curves and are affected similarly by experimental manipulations (see Biferno and Dawson 1977). Fourth, CRs are strongly influenced by language and can be both increased and decreased by instructions regarding CS–US relationships (see Grings et al. 1973; Lovibond 2003). Finally, there is evidence that CRs in more complex learning manipulations such as retrospective revaluation depend on reasoning processes (see Mitchell and Lovibond 2002). The latter two findings are particularly important in excluding an account in which conscious knowledge is an outcome or epiphenomenon of a simple conditioning process. Instead, these findings suggest that CRs derive from contingency knowledge encoded at a symbolic level.

This conclusion can be, and indeed has been, disputed (see Martin and Levey 1989; Manns et al. 2002; Wiens and Öhman

E-MAIL P.Lovibond@unsw.edu.au; FAX 61-2-9385-1193.
Article and publication are at http://www.learnmem.org/cgi/doi/10.1101/lm.79604.

2002). However, the research findings summarized above make it difficult to defend a strong version of the traditional conceptualization of conditioning as automatic, unconscious, and compartmentalized from higher-order cognition. Furthermore, similar conclusions have been reached from reviews of related forms of learning in humans, including instrumental conditioning (see Brewer 1974; Williams and Roberts 1988) and implicit learning (Shanks and St. John 1994). It seems worthwhile, then, to consider the implications of the cognitive model for how we conceptualize extinction. The remainder of this article will focus on Pavlovian conditioning, but the arguments are also likely to apply to other forms of associative learning.

Extinction in Human Pavlovian Conditioning

Human conditioning research suggests that a single mechanism governs acquisition. That single mechanism requires attentional resources, and yields associative knowledge that is represented in a propositional form, such that it can make contact with language. When the CS is encountered, it retrieves the CS–US contingency from long-term memory and thereby generates a state of expectancy of the US, which automatically triggers innate anticipatory behaviors appropriate to the US (i.e., CRs). According to this conceptualization, extinction must operate by changing associative knowledge. That is, exposure to disconfirming experiences (CS without US) must lead to revisions of the knowledge stored in long-term memory such that behavior is controlled by this updated knowledge. If US expectancy is the critical proximal basis for performance, then loss of expectancy will track extinction of the CR. What evidence is there for such a mechanism in extinction learning in humans?

Laboratory Research

Many experiments have confirmed that reductions in US expectancy track reductions in CRs during the course of extinction (see Biferno and Dawson 1977; Lipp and Edwards 2002). If contingency judgments are obtained after acquisition and then again after extinction, they demonstrate a corresponding reduction in the strength of causal belief that the US will follow the CS (see Schell et al. 1991). Similar to acquisition, extinction can be enhanced by verbal instructions. For example, "instructed extinction," in which participants are informed after acquisition that the CS will no longer be followed by the US, typically leads to an immediate reduction in the CR (see Grings et al. 1973). Failure of instructions to completely eliminate CRs can be understood in terms of factors related to the degree of confidence participants place in the instructions (Fuhrer and Baer 1980; Dawson and Schell 1985).

The cognitive model predicts that extinction will only occur when experiences contradict existing contingency beliefs—if the belief that the CS leads to the US is protected from disconfirmation, no extinction will occur. We recently provided direct evidence in support of this prediction by using the procedure of protection from extinction (Lovibond et al. 2000). Two CSs, C and D, were established as predictors of electric shock. Stimulus D was extinguished by presenting it alone, without shock. Stimulus C was also presented without shock, but it was accompanied during the extinction trials by a stimulus K that had previously been established as a safety signal, in other words as a predictor of no shock. After these extinction trials, C and D were both presented alone in a test phase (Fig. 1). A and B were control stimuli that had been consistently paired with shock and no shock, respectively. By comparison to these stimuli, it can be seen that presentation of stimulus D without shock had led to a substantial reduction in both shock expectancy (Fig. 1 left panel) and skin conductance CRs during D (Fig. 1 right panel). By contrast, stimulus C showed little extinction: The presence of K had

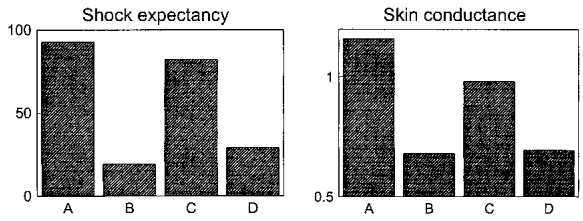

Figure 1 Mean online shock expectancy ratings (*left*) and mean change in log skin conductance level (*right*) during test trials from a protection from extinction procedure (Lovibond et al. 2000). Stimulus A was established as a consistent predictor of shock and stimulus B as a consistent predictor of no shock. Stimuli C and D were initially paired with shock and then extinguished. During extinction D was presented alone, whereas C was accompanied by a safety signal K. In the test phase data shown, C and D were both tested alone. The data show a return of responding to stimulus C, indicating that the presence of K had protected C from complete extinction. Reprinted from *Behavior Research & Therapy*, Vol. 38, Lovibond et al., "Protection from extinction in human fear conditit," pp. 967–983, © 2000, with permission from Elsevier.

"protected" it from extinction. Within the cognitive model, the absence of shock was attributed to stimulus K and, hence, the belief that C predicted shock remained intact.

The findings described above demonstrate a strong correspondence between CRs and beliefs/expectancies during extinction, consistent with the view that a single learning mechanism gives rise to both. However, if a single strong dissociation could be demonstrated between a CR measure and a self-report measure, it would provide support for the idea that there may be more than one associative mechanism. It is therefore important to examine carefully reports of such dissociations in the literature.

One such report involves the effect of the CS on learning and extinction. Seligman (1971) postulated that associations between certain fear-relevant stimuli and aversive outcomes were biologically prepared and, hence, were easy to learn and difficult to extinguish. Specifically, he suggested that phobic stimuli such as insects, snakes, and heights were prepared to enter easily into associations with aversive outcomes, and therefore individuals could show strong fear conditioning after a single learning trial. Seligman also speculated that prepared learning might be more primitive and less cognitive than nonprepared learning. In a substantial program of research over the past 30 years, Öhman and his colleagues (see Öhman et al. 1975; Öhman and Soares 1998) have subjected the notion of prepared learning to empirical examination in the laboratory. By using an autonomic conditioning procedure, Öhman (e.g., Öhman et al. 1975) found that one prediction of preparedness theory in particular could be consistently demonstrated. He observed that when fear-relevant stimuli (pictures of snakes or spiders) were used as CSs, conditioned responding took longer to extinguish than when fear-irrelevant stimuli (pictures of flowers and mushrooms) were used. This resistance to extinction effect is consistent with Seligman's (1971) idea that prepared learning might involve a different mechanism.

However, the laboratory research with fear-relevant stimuli does not provide evidence for a dissociation between CRs and self-report measures during extinction. For example, Dawson et al. (1986) replicated Öhman's basic procedure with the addition of an online measure of US expectancy. They were able to demonstrate the same resistance to extinction effect as Öhman et al. (1975) had reported, not only on the skin conductance measure but also on the expectancy measure. Similar results have been reported by Davey (1992). Initial reports that extinction of fear-relevant stimuli may be less subject to instructional effects, such as instructed extinction (Hugdahl and Öhman 1977), have not

been consistently observed (see McNally 1987; Lipp and Edwards 2002). Recent work by Öhman and colleagues using backwardly masked CS presentations has confirmed that fear-relevant stimuli behave differently from fear-irrelevant stimuli, but again self-reported US expectancy tracks the autonomic CR measure (Öhman and Soares 1998).

Furthermore, the factor that distinguishes fear-relevant from fear-irrelevant stimuli during extinction may not be associative. Evidence from my laboratory supports an alternative view that fear-relevant stimuli have an innate prepotency to elicit fear responses that may be sensitized by appropriate environmental conditions such as a separate source of arousal or anxiety (Lovibond et al. 1994; see also Menzies and Clarke 1995). According to this view, the apparent resistance to extinction of fear-relevant stimuli arises from selective sensitization of innate fear responses, superimposed on normal extinction of associative learning (Lovibond et al. 1993). Thus, conditioning and extinction with fear-relevant stimuli can be accounted for without accepting Seligman's (1971) claim that fear-relevant stimuli engage a separate, noncognitive learning mechanism.

A second extinction finding that has been interpreted as supporting multiple learning processes involves differential rates of extinction of CRs and US expectancy. In differential autonomic conditioning studies (see Dawson et al. 1986; Schell et al. 1991), differences in US expectancy ratings to CS+ and CS− sometimes disappear earlier during extinction than do differences in skin conductance responding. Thus, there is a period during the course of extinction when participants are indicating that their US expectancy is zero while still showing differential responding to CS+, supporting the idea that the skin conductance CRS might arise from a separate mechanism from expectancy. However, this single dissociation does not require the postulation of multiple processes, because it could be due to different sensitivities and/or floor effects of the two measures. Although the phenomenon deserves further empirical study with more sensitive measures, it does not in itself provide strong support for the idea of a separate low-level conditioning system (Lovibond and Shanks 2002).

Finally, several researchers (Mandel and Bridger 1967; Clark and Squire 1998) have suggested that strong CS–US contiguity (short CS–US intervals, overlap between CS and US) encourages a simpler, less cognitive type of conditioning than does weak CS–US contiguity (long CS–US intervals, trace interval between CS and US). I have previously criticized evidence from eyeblink conditioning that has been claimed to support this conclusion (Lovibond and Shanks 2002; Shanks and Lovibond 2002). In the case of autonomic conditioning, Schell et al. (1991) directly tested the impact of CS–US interval on extinction and forgetting of differential conditioning. They found that conditioning with a 0.5-sec CS–US interval was no more resistant to extinction, forgetting or instruction than conditioning with an 8-sec CS–US interval. Unaware participants, whether they had never learned or had forgotten the stimulus contingencies, showed the typical pattern of a lack of differential skin conductance responding. Thus, there is no compelling evidence that strong CS–US contiguity engages a more primitive, unconscious form of learning.

Clinical Research

Extinction-like procedures are used in the treatment of a number of disorders. For example, exposure therapy is commonly used for anxiety disorders. As in the laboratory, reductions in anxiety during exposure are accompanied by corresponding changes in threat beliefs and expectancies of harm (see Reiss 1991; Foa et al. 1996). In so-called cognitive behavioral therapy (CBT), clinicians take advantage of the synergy between behavioral and verbal procedures to draw the attention of patients to contradictions between their dysfunctional beliefs and their current experiences (see Zinbarg 1990; Lovibond 1993). For example, they combine direct demonstration of the lack of danger associated with feared stimuli (e.g., arousal symptoms in the case of panic disorder) with information and alternative explanations (e.g., hyperventilation, stress).

Interestingly, cognitive therapists have developed models of threat beliefs and their treatment that are conceptually highly similar to contemporary associative accounts. For example, Salkovskis and his colleagues (see Salkovskis 1991; Wells et al. 1995) have drawn attention to the problem of "safety behaviors" in exposure therapy for anxiety disorders. These investigators propose that anxious patients hold irrational threat beliefs; that is, they have exaggerated beliefs concerning the probability and/or cost of harmful outcomes that might occur in a given situation. Exposure therapy normally acts to disconfirm these threat beliefs because the patient is exposed to the feared situation without any harm actually occurring. However, if the patient performs safety behaviors during exposure, these behaviors act to prevent the disconfirmation that would normally occur by providing an alternative explanation for the absence of harm, thus leaving the original threat beliefs unaltered. This analysis is functionally identical to the account of protection from extinction in the laboratory outlined earlier.

Cognitive Mechanisms of Extinction

Both the laboratory and clinical evidence reviewed above demonstrate a strong correspondence between CRs and beliefs/expectancies during extinction. In both cases, the effectiveness of verbal information further supports the cognitive model. According to such a model, then, how does extinction occur? Most of the evidence relevant to this question comes from studies motivated by the testing of associative models, as there is little research directly testing the cognitive model. Nonetheless, the results of these studies can be interpreted in terms of their implications for the cognitive model. For example, both animal and human research suggests that new learning (including extinction) is initiated when the learner detects a discrepancy detected between what is expected and what actually occurs on a given learning occasion. Although Rescorla and Wagner (1972) used such a description metaphorically, the cognitive model asserts that human participants base their expectancies on their current associative knowledge, derived both from prior experience and other sources such as language, and are able to report on both their knowledge and their expectancies. This prediction is consistent with the evidence from human conditioning reviewed above, but has not been directly tested.

As noted earlier, the violation of expectancies that occurs in extinction gives rise to changes in contingency beliefs stored in long-term memory. But rather than destroying the original contingency knowledge, both animal and human research suggests that extinction establishes new learning that overrides the original learned associations. For example, Hermans, Vansteenwegen, and colleagues (Hermans et al. 2004; Vansteenwegen et al. 2004; Vervliet et al. 2004) have recently demonstrated recovery of learned fear after extinction both by a context change (renewal) and by representation of the US (reinstatement). Mineka et al. (1999) have demonstrated renewal of fear in phobic patients after exposure treatment. These findings directly parallel the results obtained from animal studies in which context changes, US memory manipulations, and reconditioning all point to the survival of the original learning (Rescorla 1996; Pearce and Bouton 2000). It appears that extinction generates new knowledge (CS leads to no outcome) that competes with the original knowledge (CS leads to US), with behavior determined by the knowledge

that is most strongly activated at a particular point in time. Alternatively, extinction may be incorporated as an exception to the original rule that is to some extent specific to the circumstances of extinction (Bouton 1994).

In humans, evidence for the preservation of earlier contingencies can be obtained purely through instructional manipulations. Collins and Shanks (2002) found that if participants were asked to make associative judgments frequently, they based their judgments on the most recently experienced trials; but if they were only asked to make a single judgment at the end of the experiment, they based their judgment on the aggregate contingency over the whole set of trials. This result suggests that participants encode changes in contingencies over time and can report flexibly on recent or past contingencies on the basis of verbal instruction. It is difficult to see how a simple associative model, in which learning is encoded only in terms of the strength of excitatory and inhibitory connections, could account for these data. The Collins and Shanks (2002) finding also helps explain why the assessment of contingency awareness after an intervening extinction phase often yields inconsistent results, whereas assessment of awareness immediately after acquisition typically yields a strong relationship with degree of conditioning (Dawson and Reardon 1973). Presumably, participants asked to indicate the association between the CS and US after extinction have two competing memories—one of CS–US pairings from acquisition, and one of CS alone presentations from extinction—and have to choose between them or combine them in some way.

The cognitive model also allows for the possibility of extinction-like effects through methods other than direct exposure to the target CS. First, extinction can be achieved by provision of verbal information, as in the laboratory procedure of instructed extinction. Second, other experiences such as observation could lead both to acquisition and extinction of associative knowledge. Finally, extinction may be achieved through the training of competing CSs. In the procedure of backward blocking, a compound of two CSs, say A and B, is initially paired with the US. In a subsequent phase, one of the CSs, say A, is paired consistently with the US. When responding to the other stimulus B is assessed, it is revealed that its ability to evoke a CR (and also a US expectancy) has been weakened by the intervening conditioning of A: a retrospective revaluation effect. Importantly, this effect is heavily influenced by training conditions that permit backward blocking as a logical inference (Mitchell and Lovibond 2002; see also De Houwer et al. 2002; Lovibond et al. 2003). This research provides strong evidence for reasoning processes in human conditioning and indicates that associative or causal beliefs can be altered by methods other than direct exposure to the CS and US.

There are some parallels between the account described above and the sorts of models developed by clinical researchers to account for changes in beliefs during treatment. Cognitive models of treatment point to the establishment of new functional schema (belief systems), rather than the destruction of prior dysfunctional schema, in order to account for the occurrence of relapse after treatment that is a feature of many clinical disorders (Beck 1996; Foa and Kozak 1986). For example, exposure therapy and corrective information are thought to help establish normative threat appraisal in anxious patients, but a single crisis can return the patient to pretreatment levels both of anxiety and threat appraisal. The notion of reactivation of earlier maladaptive schemata is very similar to the cognitive account of renewal and reinstatement after extinction in the laboratory.

A second parallel between a cognitive model of extinction and clinical models of treatment is in reattribution training. Clinicians often encourage patients to generate alternative, more adaptive explanations for past events that cannot be undone, such as failure or trauma. This strategy is directly analogous to retrospective revaluation procedures such as the example of backward blocking described earlier. By establishing the partner CS as a reliable predictor of the US, conditioning to the target CS can be reduced. This procedure is similar to the tactic of including a novel "scapegoat" CS to overshadow taste aversion learning to favorite foods during chemotherapy (see Bernstein and Webster 1985), except that it can be implemented after the conditioning episode. Furthermore, I have recently demonstrated that retrospective changes in human autonomic conditioning can be achieved not only by directly conditioning the partner CS but also by presenting verbal information about that CS (Lovibond 2003). This finding not only provides laboratory support for the clinical practice of verbal reattribution training but demonstrates that knowledge generated by conditioning must be represented in propositional form, such that it can make contact with verbal information in order to generate a retrospective inference.

Finally, it is worth noting that the analysis presented above is based on traditional Pavlovian conditioning procedures in which the CS precedes and signals the occurrence of a motivationally significant US. Under these conditions, both learning and performance appear to involve expectancy processes. However, it has been argued that associations can also be learned by a simpler system that merely associates two elements together. So-called evaluative conditioning has been proposed to differ from expectancy-based learning in three ways: First, it is said to occur independently of conscious awareness of the stimulus relationships; second, it gives rise to affective reactions of a like/dislike nature; and third (and most importantly for the present paper), it is said to be impervious to extinction once it has been established (for a review, see De Houwer et al. 2001). The distinction is a contentious one, and alternative accounts of evaluative conditioning experiments have been proposed (see Lovibond and Shanks 2002; Lipp et al. 2003). Nonetheless, even if the idea of a separate learning mechanism is not supported, it is an intriguing possibility that performance of some emotional responses might be achieved simply by activation of the memory or representation of a stimulus, without active anticipation of the actual occurrence of that stimulus. Such responses might differ from expectancy-based responses in their sensitivity to extinction procedures.

Implications for Neuroscience

A great deal of research on the neural basis of conditioning has been conducted with animals. It may be thought that evidence for the involvement of cognitive processes in human conditioning threatens the relevance of animal research. However, the striking empirical similarities between animal and human conditioning strongly support continuity of the underlying mechanisms (Lovibond and Shanks 2002). Rather than accepting the traditional view that conditioning is a low-level process in both animals and humans, it is more parsimonious to assume that laboratory animals possess a precursor to the human cognitive system, one that is capable of representing certain aspects of the environment. In other words, the mechanisms that allow conditioning in animals evolved into those that allow higher-level cognition and language in humans. This view in fact provides a stronger justification for animal research than does the traditional model, because it postulates that conditioning in animals will be relevant not only to conditioning in humans but also to other high-level cognitive processes.

The implications of the cognitive model for neuroscience research are very different from those of the traditional model. The primary implication concerns the way in which associative information is stored and used. The traditional model suggests that conditioning and extinction can both be understood in

terms of associative connections between "nodes" corresponding to the CS and US/UR. These connections may be excitatory or inhibitory, but they are constrained to a single property, the regulation of activation. As such, they are functionally the same as the traditional conception of reflexes. This view encourages neuroscientists to look for direct neural connections between the brain systems responsible for perception of the CS and the US, and/or between the CS and the systems responsible for emotional or motor responding. The cognitive model, by contrast, postulates that learning results in a symbolic internal representation of the relationship between the CS and US that includes information such as temporal sequence and timing. Such a representation is unlikely to be coded in terms of a direct neural connection at the perceptual or motor level. Furthermore, in the cognitive model performance typically derives not from the direct activation of the US representation (as though the US itself had been presented), but from a distinct state of expectancy of the US. Expectancy processes may give rise to anticipatory species-specific behavior appropriate to the motivational category of the US (e.g., anxiety in the case of harmful USs or craving in the case of appetitive USs), rather than or in addition to responses specific to the particular US used (see Konorski 1967). Although expectancy appears to be central to many responses such as anxiety and craving, it is possible that other responses such as evaluative (like/dislike) reactions require only activation of the US representation or its motivational system. An important task for neuroscience is to delineate the neural processes underlying memory and expectancy.

Similar to acquisition, extinction depends on surprise, as demonstrated by the phenomenon of protection from extinction. However, the surprise that is generated when an expected event fails to occur is more focused than when an unexpected event does occur. Extinction is intrinsically more complex than is acquisition because it depends on retrieval of specific prior learning from long-term memory. Furthermore, as already noted, extinction appears to preserve the original learning and add new learning that suppresses the original learned behavior. This new learning is often conditionalized on features such as context or time and can be disrupted by methods that do not disrupt initial learning. Thus, it may reasonably be predicted that extinction involves more complex processes than acquisition and, hence, may depend on additional brain regions and/or be affected differently by pharmacological manipulations.

In the case of human neuroscience research, the primary implication of the cognitive model is that there is a single learning process to be understood, rather than two separate processes (Lovibond and Shanks 2002). That single process will include systems responsible for language, reasoning, and conscious representation of contingencies, as well as the expectation of future events. It is important to note that the cognitive view does not deny the role of low-level, unconscious processes in conditioning. Such processes as attention, perception, and retrieval of information from long-term memory must be involved prior to the point of conscious representation of associative relationships. What the cognitive model does deny, however, is that low-level systems alone are capable of generating CRs. That is, low-level processes are necessary for higher-level processes such as those associated with conscious awareness, but they are not by themselves sufficient to yield learning. They enable higher-level cognition rather than competing with it. Thus, neuroscientists should be looking for an integrated system that includes both low-level and higher-level processes rather than multiple competing systems each capable of generating behavior. Apparent dissociations need to be examined carefully to see whether they demand postulation of multiple processes rather than a single integrated process (Dunn and Kirsner 2003).

Conclusions

I have reviewed evidence for a cognitive model of extinction that encompasses both animal and human evidence. Future research needs to test novel predictions of the cognitive model and flesh out aspects that are presently descriptive rather than explanatory. Clinical psychologists have already developed procedures consistent with the cognitive model of extinction, and these procedures may be further refined by taking advantage of laboratory models. The challenge for neuroscience is to map the known functional properties of conditioning and extinction on to neural systems. This task will require moving beyond the traditional view that cortical regions are associated with higher-level cognition, whereas subcortical structures are associated with phylogenetically older, more primitive noncognitive systems, operating in a competitive fashion. Conditioning and extinction are more likely to involve multiple brain processes operating in an integrated, cooperative fashion.

ACKNOWLEDGMENTS

Preparation of this manuscript was supported by grants A10007156 and DP0346379 from the Australian Research Council. I would like to thank Chris Mitchell for his helpful comments on the manuscript.

REFERENCES

Beck, A.T. 1996. Beyond belief: A theory of modes, personality, and psychopathology. In *Frontiers of cognitive therapy* (ed. P.M. Salkovskis), pp. 1–25. Guilford Press, New York.

Bernstein, L. and Webster, M.M. 1985. Learned food aversions: A consequence of cancer chemotherapy. In *Cancer, nutrition and eating behavior: A biobehavioral perspective* (eds. T.G. Burish et al.), pp. 103–116. Erlbaum, Hillsdale, NJ.

Biferno, M.A. and Dawson, M.E. 1977. The onset of contingency awareness and electrodermal classical conditioning: An analysis of temporal relationships during acquisition and extinction. *Psychophysiology* **14:** 164–171.

Bouton, M.E. 1994. Context, ambiguity, and classical conditioning. *Curr. Dir. Psychol. Sci.* **3:** 49–53.

Brewer, W.F. 1974. There is no convincing evidence for operant or classical conditioning in adult humans. In *Cognition and the symbolic processes* (eds. W.B. Weimer and D.S. Palemo), pp. 1–42. Erlbaum, Hillsdale, NJ.

Clark, R.E. and Squire, L.R. 1998. Classical conditioning and brain systems: The role of awareness. *Science* **280:** 77–81.

Collins, D.J. and Shanks, D.R. 2002. Momentary and integrative response strategies in causal judgment. *Mem. Cognit.* **30:** 1138–1147.

Davey, G.C.L. 1987. An integration of human and animal models of Pavlovian conditioning: Associations, cognitions, and attributions. In *Cognitive processes and Pavlovian conditioning in humans* (ed. G.C.L. Davey), pp. 83–114. Wiley, Chichester, UK.

———. 1992. An expectancy model of laboratory preparedness effects. *J. Exp. Psychol.* **121:** 24–40.

Dawson, M.E. 1970. Cognition and conditioning: Effects of masking the CS–UCS contingency on human GSR classical conditioning. *J. Exp. Psychol.* **85:** 389–396.

Dawson, M.E. and Reardon, P. 1973. Construct validity of recall and recognition postconditioning measures of awareness. *J. Exp. Psychol.* **98:** 308–315.

Dawson, M.E. and Schell, A.M. 1985. Information processing and human autonomic classical conditioning. In *Advances in psychophysiology*, Vol. 1 (eds. P.K. Ackles et al.), pp. 89–165. JAI Press, Greenwich, CT.

Dawson, M.E., Schell, A.M., and Banis, H.T. 1986. Greater resistance to extinction of electrodermal responses conditioned to potentially phobic CSs: A noncognitive process? *Psychophysiology* **23:** 552–561.

De Houwer, J., Thomas, S., and Baeyens, F. 2001. Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychol. Bull.* **127:** 853–869.

De Houwer, J., Beckers, T. and Glautier, S. 2002. Outcome and cue properties modulate blocking. *Q. J. Exp. Psychol. A* **55:** 965–985.

Dunn, J.C. and Kirsner, K. 2003. What can we infer from double dissociations? *Cortex* **39:** 1–7.

Foa, E.B. and Kozak, M.J. 1986. Emotional processing of fear: Exposure to corrective information. *Psychol. Bull.* **99:** 20–35.

Foa, E.B., Franklin, M.E., Perry, K.J., and Herbert, J.D. 1996. Cognitive

biases in generalized social phobia. *J. Abnormal Psychol.* **105:** 433–439.
Fuhrer, M.J. and Baer, P.E. 1980. Cognitive factors and CS–UCS interval effects in the differential conditioning and extinction of skin conductance responses. *Biol. Psychol.* **10:** 283–298.
Grings, W.W., Schell, A.M., and Carey, C.A. 1973. Verbal control of an autonomic response in a cue reversal situation. *J. Exp. Psychol.* **99:** 215–221.
Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., and Eelen, P. 2004. Reinstatement of fear responses in human aversive conditioning. *Behav. Res. Ther.* (in press).
Hugdahl, K. and Öhman, A. 1977. Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli. *J. Exp. Psychol. Human Learn. Mem.* **3:** 608–618.
Konorski, J. 1967. *Integrative activity of the brain.* University of Chicago Press, Chicago.
Lipp, O.V. and Edwards, M.S. 2002. Effect of instructed extinction on verbal and autonomic indices of Pavlovian learning with fear-relevant and fear-irrelevant conditional stimuli. *J. Psychophysiol.* **16:** 176–186.
Lipp, O.V., Oughton, N., and LeLievre, J. 2003. Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learn. Motivation* **34:** 219–239.
Lovibond, P.F. 1993. Conditioning and cognitive-behaviour therapy. *Behav. Change* **10:** 119–130.
———. 2003. Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *J. Exp. Psychol. Learn. Mem. Cognit.* **29:** 97–106.
Lovibond, P.F. and Shanks, D.R. 2002. The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *J. Exp. Psychol. Anim. Behav. Proc.* **28:** 3–31.
Lovibond, P.F., Siddle, D.A.T., and Bond, N. 1993. Resistance to extinction of fear-relevant stimuli: Preparedness or selective sensitization? *J. Exp. Psychol.* **122:** 449–461.
Lovibond, P.F., Hanna, S.K., Siddle, D.A.T., and Bond, N.W. 1994. Electrodermal and subjective reactions to fear-relevant stimuli under threat of shock. *Aus. J. Psychol.* **46:** 73–80.
Lovibond, P.F., Davis, N.R., and O'Flaherty, A.S. 2000. Protection from extinction in human fear conditioning. *Behav. Res. Ther.* **38:** 967–983.
Lovibond, P.F., Been, S-L., Mitchell, C.J., Bouton, M.E., and Frohart, R. 2003. Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Mem. Cognit.* **31:** 133–142.
Mandel, I.J. and Bridger, W.H. 1967. Interaction between instructions and ISI in conditioning and extinction of the GSR. *J. Exp. Psychol.* **74:** 36–43.
Manns, J.R., Clark, R.E., and Squire, L.R. 2002. Standard delay eyeblink classical conditioning is independent of awareness. *J. Exp. Psychol. Anim. Behav. Proc.* **28:** 32–37.
Martin, I. and Levey, A.B. 1989. Propositional knowledge and mere responding. *Biol. Psychol.* **28:** 149–155.
McNally, R.J. 1987. Preparedness and phobias. *Psychol. Bull.* **101:** 283–303.
Menzies, R.G. and Clarke, J.C. 1995. The etiology of phobias: A nonassociative account. *Clin. Psychol. Rev.* **15:** 23–48.
Mineka, S., Mystkowski, J.L., Hladek, D., and Rodriguez, B.I. 1999. The effects of changing contexts on return of fear following exposure therapy for spider fear. *J. Consult. Clin. Psychol.* **67:** 599–604.
Mitchell, C.J. and Lovibond, P.F. 2002. Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *Q. J. Exp. Psychol.* **55B:** 311–329.
Öhman, A. and Soares, J.J.F. 1998. Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. *J. Exp. Psychol.* **127:** 69–82.
Öhman, A., Erixon, G., and Lofberg, I. 1975. Phobias and preparedness: Phobic versus neutral pictures as conditioned stimuli for human autonomic responses. *J. Abnorm. Psychol.* **84:** 41–45.
Pearce, J.M and Bouton, M.E. 2000. Theories of associative learning in animals. *Ann. Rev. Psychol.* **52:** 111–139.
Razran, G. 1955. Conditioning and perception. *Psychol. Rev.* **62:** 83–95.
Reiss, S. 1991. Expectancy model of fear, anxiety, and panic. *Clin. Psychol. Rev.* **11:** 141–153.
Rescorla, R.A. 1996. Preservation of Pavlovian associations through extinction. *Q. J. Exp. Psychol.* **49B:** 245–258.
Rescorla, R.A. and Wagner, A.R. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current research and theory* (eds. A.H. Black and W.F. Prokasy), pp. 64–99. Appleton-Century-Crofts, New York.
Salkovskis, P.M. 1991. The importance of behaviour in the maintenance of anxiety and panic: A cognitive account. *Behav. Psychother.* **19:** 6–19.
Schell, A.M., Dawson, M.E., and Marinkovic, K. 1991. Effects of potentially phobic conditioned stimuli on retention, reconditioning, and extinction of the conditioned skin conductance response. *Psychophysiology* **28:** 140–153.
Seligman, M.E.P. 1971. Phobias and preparedness. *Behav. Ther.* **2:** 307–320.
Shanks, D.R. and Lovibond, P.F. 2002. Autonomic and eyeblink conditioning are closely related to contingency awareness: Reply to Wiens and Öhman (2002) and Manns et al (2002). *J. Exp. Psychol. Anim. Behav. Proc.* **28:** 38–42.
Shanks, D.R. and St. John, M.F. 1994. Characteristics of dissociable human learning systems. *Behav. Brain Sci.* **17:** 367–447.
Squire, L.R. 1994. Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In *Memory systems* (eds. D.L. Schacter and E. Tulving), pp. 203–231. MIT Press, Cambridge, MA.
Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., and Eelen, P. 2004. Return of fear in a human differential conditioning paradigm caused by a return to the original acquisition context. *Behav. Res. Ther.* (in press).
Vervliet, B., Vansteenwegen, D., Baeyens, F., Hermans, D., and Eelen, P. 2004. Differential impact of stimulus change after acquisition versus extinction treatment in human fear conditioning. *Behav. Res. Ther.* (in press).
Wells, A., Clark, D.M., Salkovskis, P., Ludgate, J., Hackmann, A., and Gelder, M. 1995. Social phobia: The role of in-situation safety behaviours in maintaining anxiety and negative beliefs. *Behav. Ther.* **26:** 153–161.
Wiens, S. and Öhman, A. 2002. Unawareness is more than a chance event: Comment on Lovibond and Shanks (2002). *J. Exp. Psychol. Anim. Behav. Proc.* **28:** 27–31.
Williams, R.J and Roberts, L.E. 1988. Relation of learned heart rate control to self-report in different task environments. *Psychophysiology* **25:** 354–365.
Zinbarg, R.E. 1990. Animal research and behavior therapy, part I: Behavior therapy is not what you think it is. *The Behavior Therapist* **13:** 171–175.