

Extinction in Human Fear Conditioning

Dirk Hermans, Michelle G. Craske, Susan Mineka, and Peter F. Lovibond

Although most extinction research is conducted in animal laboratories, the study of extinction learning in human fear conditioning has gained increasing attention over the last decade. The most important findings from human fear extinction are reviewed in this article. Specifically, we review experimental investigations of the impact of conditioned inhibitors, conditioned exciters, context renewal, and reinstatement on fear extinction in human samples. We discuss data from laboratory studies of the extinction of aversively conditioned stimuli, as well as results from experimental clinical work with fearful or anxious individuals. We present directions for future research, in particular the need for further investigation of differences between animal and human conditioning outcomes, and research examining the role of both automatic and higher-order cognitive processes in human conditioning and extinction.

Key Words: Associative learning, fear conditioning, extinction, renewal, reinstatement, exposure

Extinction can be described as a procedure, as a result, or as a process. As a procedure, extinction refers to the repeated presentation of a conditioned stimulus (CS) in the absence of the unconditioned stimulus (US) with which it was previously paired. Following a series of CS-only presentations, the conditioned response (CR) that was established during the preceding acquisition phase gradually diminishes. For instance, during fear conditioning, the contingent presentation of a tone CS and an aversive shock (US) will alter the response to the former stimulus; i.e., the CS will come to evoke fear responses when tested in absence of the US. Subsequent nonreinforced presentations of the tone CS will lead to a reduction in these acquired fear CRs.

Although there are many aspects of extinction that are still not fully understood (Rescorla 2001), the past 15 years have seen significant developments in the conceptualization of the processes of extinction. For instance, insights into the anatomical and biochemical basis of the extinction of conditioned fear have been growing steadily (for an overview, see Barad, in press; Myers and Davis 2002). Moreover, some of these insights have directly informed possible future treatments for anxiety disorders (Ressler et al 2004; Walker et al 2002).

Another development that has important theoretical and clinical implications is the growing recognition that the diminished (or even vanished) responding that is observed after extinction, whether measured behaviorally or psychophysiological, does not reflect "unlearning" at the process level (e.g., Bouton 1988; Delamater 2004). Within an associative approach to Pavlovian conditioning, this means that extinction does not (completely) destroy the original CS-US association. Empirical support comes from a large amount of animal work showing that under the right conditions, extinguished conditioned responses can be recovered (for an overview, see Bouton 1994, 2004; Falls 1998). Examples of such conditions are the mere passage of time (spontaneous recovery) (Pavlov 1927; Rescorla 2004), context change after extinction (renewal) (e.g., Bouton and Ricker 1994), or the presentation of postextinction US-only trials (reinstatement)

(e.g., Rescorla and Heth 1975). Whereas these phenomena support the view that extinction does not completely destroy the CS-US association, more direct tests of the CS-US association in animals actually suggest that extinction leaves the CS-US association intact (Delamater 1996; Rescorla 1996).

These developments in extinction research have been largely based on work with laboratory rodents. Hence, it remains an open question as to whether these findings can be generalized to humans. Clearly, there are large differences in brain anatomy between rats on the one hand and humans on the other. Given the often highly specialized functions of specific brain regions, there are no a priori reasons to assume that identical processes are involved in extinction in human and nonhuman animals. Although there is now a fair amount of work concerning the neurobiological basis of fear acquisition in humans (e.g., Phelps et al 2001), the study of the neural mechanisms of extinction learning in humans has received less attention (e.g., Knight et al 2004; LaBar et al 1998; Phelps et al 2004). For an overview of this literature, we refer to Rauch et al (in press).

Another reason why animal findings might not readily translate to human samples is that humans are endowed with a cognitive system that is capable of a symbolic and propositional analysis of conditioning experiences (see De Houwer et al 2005; Lovibond and Shanks 2002). The extent to which higher level propositional processes are involved in associative learning in general, and in fear conditioning and extinction more specifically, has been a matter of debate (see Beaver et al 2005; Carter et al 2003; Clark and Squire 1998; Lovibond 2003, 2004; Öhman and Mineka 2001).

The translation of animal extinction research to human samples has received more attention lately due to the clinical relevance of the topic. That is, extinction of conditioned fear can be viewed as a laboratory analog for the study of exposure treatment for anxiety disorders (Bouton et al 2001; Davey 1997; Eelen et al 2001; Mineka 1985). Hence, insight into the mechanisms that are responsible for fear extinction may have clinical significance. In addition, knowledge of the conditions that facilitate or hamper extinction learning may help to sharpen exposure treatments in such a way that treatment outcome can be maximized in the short run (therapy efficacy) as well as in the long run (relapse prevention). In this article, we review recent experimental work on extinction of human fear learning. In general, two different approaches have been employed in this context. The first approach mainly uses Pavlovian preparations to study procedures and processes of acquisition and extinction in the laboratory. These experiments typically employ a differential aversive conditioning procedure with nonclinical groups. The second approach takes these observations one step further by translating the extinction principles to exposure therapy

From the Katholieke Universiteit Leuven (DH), Leuven, Belgium; University of California (MGC), Los Angeles, California; Northwestern University (SM), Evanston, Illinois; and University of New South Wales (PFL), Sydney, Australia.

Address reprint requests to Dirk Hermans, Department of Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium; E-mail: Dirk.Hermans@psy.kuleuven.ac.be.

Received June 2, 2005; revised October 5, 2005; accepted October 27, 2005.

protocols in clinical groups with established fears. The combination of these two methods offers a fruitful approach to study fundamental as well as clinical aspects of extinction of conditioned fear (Vansteenwegen et al, *in press*).

Experimental Laboratory Studies

Protection from Extinction and Super-Extinction

Various theoretical models of Pavlovian conditioning allow concrete predictions about the conditions that influence the process of extinction and either facilitate or hamper its behavioral outcome. Examples are the effects of trial spacing in extinction, the impact of the associative strength of contextual cues, the nature of the CS (e.g., fear-relevance), the presentation of other excitatory CSs during extinction, or the introduction of inhibitory CSs. Particularly the latter phenomenon has received great interest within the clinical field, as it is widely believed that such inhibitory “safety-signals” can have a deleterious effect on the progress of exposure. For instance, the presence of the therapist during exposure to anxiety-provoking cues can act as a conditioned inhibitor (“when the therapist is present, nothing bad will happen”). The presence of the therapist may lead to rapid symptom reduction, but a return of anxiety may occur when the client subsequently confronts the stimuli alone. In terms of the Rescorla-Wagner model (Rescorla and Wagner 1972), adding an inhibitory stimulus during extinction blocks extinction of the excitatory CS. Given that the inhibitory stimulus has a negative associative strength that cancels out the positive associative strength of the excitatory stimulus, this will lead to little or no discrepancy between what is predicted by the sum of both cues (close to zero) and what actually happens on extinction trials (zero). As a result, there is little change in associative strength of the stimuli, and in particular, the excitatory CS should retain a positive associative strength, revealed when it is tested alone. This phenomenon is known as protection from extinction and has been demonstrated in animal research using procedures of aversive Pavlovian conditioning (Soltysik et al 1983) and autoshaping (Rescorla 2003).

Employing a differential aversive conditioning procedure with colored blocks and a 500-Hz tone as CSs and .5-second shock as US, Lovibond et al (2000) (Experiment 1) tested the protection from extinction hypothesis in a group of undergraduate students.

In the acquisition phase of this study, two CSs (stimulus C and stimulus D) were contingently followed by the US and were hence established as valid predictors of this aversive stimulus. Stimulus D was subsequently extinguished by presenting it alone without shock. Stimulus C was also presented without shock, but during the extinction trials, it was accompanied by a stimulus K that had previously been established as an inhibitory stimulus (i.e., following a procedure of conditioned inhibition; A+/AK−). Following these extinction trials, stimulus C and stimulus D were presented alone in a test phase (see Figure 1). Stimulus A and stimulus B were control stimuli that had been consistently paired with shock and no shock, respectively. Compared with stimulus A, presentation of stimulus D without shock led to a substantial reduction in both shock expectancy (left panel of Figure 1) and electrodermal responses during stimulus D (right panel of Figure 1). By contrast, stimulus C showed little extinction. The presence of stimulus K had protected it from extinction.

Human studies have not yet demonstrated that the added stimulus needs to be inhibitory to interfere with extinction. In fact, it has been shown that an excitatory stimulus can have the same effect. According to the Rescorla-Wagner model, adding a

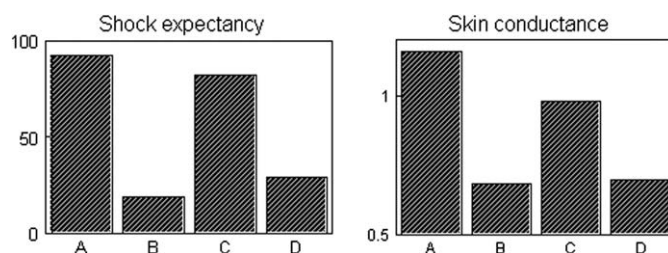


Figure 1. Mean online shock expectancy ratings (left panel) and mean change in log skin conductance level during test trials from a protection from extinction procedure (Lovibond et al 2000). Stimulus A was established as a consistent predictor of shock and stimulus B as a consistent predictor of no shock. Stimuli C and D were initially paired with shock and then extinguished. During extinction, stimulus D was presented alone, whereas stimulus C was accompanied by a safety signal K. In the test phase data shown, stimuli 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 stimulus C from complete extinction. (Reproduced with permission from Lovibond et al 2000).

conditioned exciter to the target CS during extinction should actually enhance extinction of the target CS. Although there is evidence for such super-extinction in animal studies (e.g., Rescorla 2000; Thomas and Ayres 2004), two studies in human participants have demonstrated protection from extinction by an excitatory CS (Lovibond et al 2000, Experiment 2; Vervliet et al 2005). Hence, future research should target possible differences between human and nonhuman conditioning procedures concerning the effect of nonreinforcement in the presence of another excitatory stimulus. For example, the ceiling effect imposed by ethical constraints on US magnitude in humans may disrupt the normal process of additivity that underpins the Rescorla-Wagner model (Mitchell and Lovibond 2002). Alternatively, it is possible that humans are particularly sensitive to generalization decrement or contextual changes, such that extinction is highly specific to the precise circumstances under which nonreinforcement has occurred.

Contextual and Stimulus Specificity of Fear Extinction

The studies described above show how differences in extinction procedures can affect responding to the extinguished CS later at test. As shown by the phenomenon of protection from extinction, responding that disappeared as a result of apparently successful extinction may nevertheless reappear afterward. As was already pointed out, context changes might also lead to recovery of apparently extinguished responses. The phenomenon of renewal is a good example of the context-dependency of extinction, which is most frequently observed in ABA-renewal designs. Bouton and Swartztruber (1986) demonstrated that conditioned responding is renewed when animals are tested in the original acquisition context after extinction in a different context. Renewal effects also are obtained when, after extinction in a different context, transfer of extinction is tested in a third context (ABC-renewal) and in AAB-renewal designs in which acquisition and extinction take place in an identical context but testing is executed in a new context (Bouton and Ricker 1994).

In a recent study, Vansteenwegen et al (2005) studied ABA-renewal in a human differential fear conditioning preparation. The CSs in this study were line drawings of pictorial faces. Contexts were manipulated by turning the central lighting of the experimental room on or off. An ABA-renewal group and an AAA-control group were compared. The renewal group received acquisition in one context (dark/light) and extinction in the

opposite context (light/dark) and was tested again with the CSs in the original acquisition context. In the control group, all phases were executed in the same context (which could be either light or dark, depending on the counterbalancing scheme). Although no change in differential electrodermal responding was observed when switching the context after acquisition (i.e., there was generalization of acquisition to the new context), there was a clear return of conditioned responding with the context change after extinction. No such return was observed in the control group. Retrospective US-expectancy ratings confirmed this data pattern. Thus, in line with animal research, this study demonstrated that previously extinguished fear responses can return when tested in a context different from the extinction context. Similar results were recently obtained in a study by Milad et al (2005).

In a subsequent study, Vervliet et al (2005) investigated whether recovery of extinguished fear responses is observed when stimuli rather than contexts were manipulated. In this experiment, participants learned that one of two geometrical figures (e.g., a parallelogram or a triangle) was followed by an aversive US (shock). During the subsequent extinction phase, variants of these CSs were presented (angles were slightly more sharp/blunt) for the experimental group. For the control group, the original acquisition CSs were employed. Testing involved the acquisition CSs for both groups. Differential acquisition effects generalized substantially to the extinction stimuli in the experimental group. However, similar to what is observed in ABA-renewal, participants from the experimental group hardly generalized from what was learned with the extinction stimuli to the original acquisition stimuli during final testing. These findings were later replicated in a follow-up study (Vervliet et al 2004).

To what extent the same mechanisms drive these contextual effects in humans and animals needs to be subjected to further experimental analysis (Vansteenwegen et al, in press). As already indicated, the importance of higher-order cognitive processes in humans also needs closer examination. For example, it may be possible to establish or reverse safety signals in humans through verbal instruction (e.g., Norrholm et al, unpublished data).

Reinstatement Following Fear Extinction

Another paradigm that has received attention in the context of extinction research is reinstatement. In a reinstatement procedure, a previously conditioned stimulus is first extinguished until conditioned responding disappears. Next, a few USs are administered in the absence of the CS. When the previously extinguished CS is finally re-presented, conditioned responding to the CS (partially) reappears (e.g., Rescorla and Heth 1975; Richardson et al 1999). It is assumed that the US-only presentations reinstate the CS-US association that is still present but dormant after extinction. Research has pointed to the crucial role of the context for the emergence of reinstatement effects in animals. Specifically, reinstatement effects only occur when USs are presented in the original conditioning context (for an overview and discussion, see Bouton 2004; Westbrook et al 2002). As a matter of fact, in a human reinstatement study using a single-cue paradigm, LaBar and Phelps (2005) demonstrated reinstatement effects to be context-specific.

In a series of differential fear conditioning experiments, Hermans and colleagues studied reinstatement in humans. In the first experiment, Hermans et al (2005) used pictures of neutral faces as CSs and an electrocutaneous stimulus as the US. During acquisition training, one stimulus (CS+) was contingently followed by the US, while a second stimulus (CS-) predicted the absence of the US. Subsequently, both CSs were involved in a

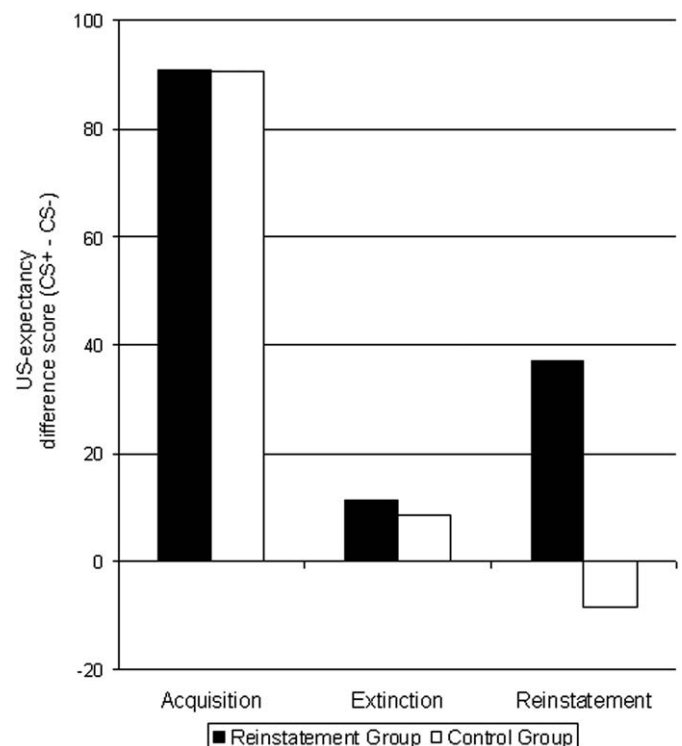


Figure 2. Differences in US-expectancy for the CS+ and CS- as a function of condition (reinstatement group, control group) and moment (postacquisition, postextinction, postreinstatement) (difference score = CS+ - CS-). (Reproduced with permission from Hermans et al 2005). US, unconditioned stimulus; CS+, stimulus that was contingently followed by US during acquisition; CS-, stimulus that was never followed by US during acquisition.

lengthy extinction phase. For half of the participants (reinstatement group), extinction training was followed by the presentation of four unpredicted USs (in the same context). For the control group, no such US-only presentations were scheduled. As can be seen from Figure 2, the groups showed equal acquisition and extinction effects for the US-expectancy ratings. Crucially, when the CSs were tested after reinstatement, there was a selective partial return of US-expectancy for the CS+ in the reinstatement group. This effect was absent in the control group. Analogous to the US-expectancy ratings, selective reinstatement was observed in the fear ratings as well. These findings were replicated in a study by Dirikx et al (unpublished data).

However, a limitation of both of these studies is that the reinstatement effects were restricted to self-report measures of fear and US-expectancy. In a study by Dirikx et al (2004), reinstatement was also demonstrated using a reaction time measure of fear that indexed allocation of attentional resources.¹ In this task, participants press a key as soon as a tone probe is

¹The reaction time task that was used in this experiment was originally developed by Dawson et al (1982) and has been successfully employed as a measure of allocation of processing resources to the CS+/CS- in studies of human Pavlovian conditioning (see also Hermans et al 2005; Lipp et al 1993). There has been a growing interest in this type of procedure. Other reaction time procedures that have been used in human fear conditioning research include the visual dot probe reaction time task (e.g., Beaver et al 2005) and the exogenous cuing paradigm (Koster et al 2004). To assess the affective meaning of the conditional stimuli, the affective priming procedure has been employed as a nonverbal measure (e.g., Hermans et al 2002).

presented. It was demonstrated that during acquisition, response latencies were significantly slowed when tones were presented during the CS+ as compared with the CS-. This difference disappeared as a result of extinction training. After reinstatement, however, a selective slowing for the CS+ (as compared with the CS-) was observed again in the reinstatement group.

Another intriguing finding in this series of reinstatement studies is that the (extinction resistant) negative valence of the CS+ (or the affective discrimination between the CS+ and the CS-) was predictive of the extent to which fear returned.² Hermans et al (2005) showed that to the extent that the CS+ had a more negative valence after extinction, there was a more significant return of fear responses. The more negative the CS+, the more recovery of the extinguished fear responses. This effect could only be attributed to the negative valence of the CS+ and was unrelated to other characteristics of the CS+ after extinction. Similar findings were reported by Dirikx et al (unpublished data) who showed that the valence of the CS+ or the difference in valence between CS+ and CS- (Dirikx et al 2004) was predictive of the amount of return of fear. Depending on the study, this predictive power was restricted to the reinstatement group (reinstatement effect) or concerned both groups as a whole (general return of fear). These results clearly show that negative stimulus valence functions as a separate source for the recovery of extinguished fear responses. A possible explanation for these findings is that insofar as fear is based on an orthogonal combination of negative valence and high arousal (e.g., Lang et al 1990), extinction might change the arousal component but not the valence component. According to that view, adding arousal through US-only presentations would reinstate fear to the extent that the CS is still negative in valence.

Based on this overview of experiments, we can conclude that a number of extinction phenomena that were previously demonstrated in animal research can be replicated in experimental laboratory studies in humans (e.g., protection from extinction, renewal, reinstatement). The fact, however, that similar procedures lead to similar effects does not imply that similar processes are at work. Further research is needed to illuminate similarities and differences at the process level. Nevertheless, these findings have clear clinical implications about what procedures could be employed to facilitate exposure and reduce return of fear.

Experimental Clinical Studies

The experimental investigation of extinction procedures in human phobic samples is immediately disadvantaged by starting from the stage of fear reduction rather than fear acquisition. Thus, the circumstances of acquisition are often unknown and there is usually no opportunity to test in the original context. Nonetheless, such studies are of great importance because of their potential relevance for enhancing treatments for anxiety disorders. Moreover, replicable effects have been observed. This section reviews experimental investigations of conditioned inhibitors, conditioned excitors, context renewal, and reinstatement effects in human phobic samples.

²In line with findings from nonaversive conditioning procedures (for an overview, see De Houwer et al 2001), it has been demonstrated that differential fear conditioning procedures can lead to an affective shift in the valence of the CS+ (which becomes a more negatively valenced stimulus). This conditioned valence seems relatively resistant to extinction (see Hermans et al 2002; Lipp and Purkis 2005).

Safety Signals and Protection from Extinction

Although clinicians are typically encouraged to wean their clients from safety signals (such as the presence of a therapist, pill bottles, etc.) during exposure therapy (e.g., Barlow and Craske 1994), the role of conditioned inhibitors during exposure therapy for phobias remains poorly studied. Investigations to date have either failed to experimentally establish safety signals as conditioned inhibitors and/or have confounded safety signals with the primary conditional/phobic stimulus.

Sloan and Telch (2002) reported that claustrophobic participants who received an exposure treatment in which they were encouraged to use safety signals reported more fear at posttest and follow-up than those encouraged to focus on their fear during exposure. In a subsequent study, Powers et al (2004) found that the perception of safety (i.e., availability of safety behaviors regardless of whether they were used) rather than use of safety was detrimental to treatment outcome, since level of fear reduction was unaffected by actual use of safety behaviors. However, in both studies, the effects of safety signal encouragement may have been attributable to distraction and purported safety signals were not established experimentally as safety signals. Moreover, the safety signals (i.e., opening a window, unlocking a door lock) constituted behaviors that in essence degraded the conditioned stimulus (i.e., enclosed situations for claustrophobic participants). Hence, they more closely represent avoidance responses rather than safety signals.

However, avoidance responses may share some functional properties with Pavlovian safety signals. Consistent with this idea, Salkovskis and his colleagues (e.g., Salkovskis 1991; Wells et al 1995) have provided evidence that “within-situation safety behaviors,” equivalent to avoidance responses, interfere with the benefits of exposure therapy. Specifically, they showed that teaching anxious clients to refrain from these behaviors leads to greater fear reduction after an exposure session. These researchers argue that clients attribute the absence of harm to their safety behaviors, rather than reducing their appraisal of the threatening stimulus. An alternative interpretation of these results derives from a protection from extinction perspective. Exposure to a phobic stimulus while using safety signals (avoidance responses/conditioned inhibitors) partially protects the phobic stimulus from extinction. Clearly, much more direct investigation is needed on the effects of safety signals and avoidance responses during exposure, especially given the very direct implications for clinical practice.

As noted earlier, two human laboratory studies have failed to show that concurrent conditioned excitors enhance extinction. In the realm of clinical phobias, inclusion of multiple excitors during exposure therapy is sometimes advocated, such as occurs when interoceptive exposure (e.g., consuming caffeine) is incorporated with exposure to feared situations (e.g., traveling by subway) for the treatment of panic disorder with agoraphobia (Barlow and Craske 1994). However, no studies have directly evaluated multiple excitors in a clinical population.

Contextual Effects on Return of Fear

By far, the majority of research to date has focused on context renewal effects in human phobic samples, wherein contexts associated with successful exposure treatment are assumed to activate nonfearful memories and nontreatment contexts are assumed to elicit fearful memories and return of fear. Studies to date have evaluated effects of contextual shifts on renewal in individuals who are highly fearful of spiders, who undergo graduated exposure therapy with a live spider using participant

modeling, conducted within one session. Outcome is assessed immediately posttreatment, and retesting occurs 1 to 2 weeks later. Fear typically is measured via self-report, behavioral avoidance, and heart rate during a behavioral approach test to a live spider.

In general, context renewal effects have been observed when participants are retested 1 to 2 weeks later in a context that differs from the treatment context (e.g., Rodriguez et al 1999; Mineka et al 1999). The effects became stronger with more distinctly different contexts and with various other improvements to methodology. Specifically, Mystkowski et al (2002) compared an outside context to the standard laboratory context. As was done in the study by Mineka et al (1999), the contexts were additionally differentiated by the sex of the experimenter and salient visual cues (i.e., color of treatment materials and therapist lab coats), and to overcome prior methodological limitations, self-report fear measures were administered in the experimental context (versus the instruction room) throughout the study. Also, this study utilized a within-subjects paradigm, similar to designs sometimes used in the animal conditioning literature (e.g., Bouton and Brooks 1993) to maximize power. Thus, 1 week later, participants were tested for return of fear in both the original treatment context and the different context in a counter-balanced order (AB or BA). Self-reported return of fear was greater when individuals confronted a previously feared stimulus in a context different than the extinction context (see Figure 3), and the effect sizes and power were much larger than in prior context studies. However, the findings were restricted to self-reported fear obtained during behavioral approach tasks, and the B-B/A condition failed to produce renewal, for reasons that are not immediately clear. In addition, these studies, in general, do not find context effects in terms of how closely participants approach the phobic stimulus. This may be due to ceiling effects, because exposure treatments effectively eliminated behavioral avoidance (Mystkowski and Mineka, *in press*).

Context-based renewal also has been demonstrated via use of videotaped presentations. Specifically, Vansteenwegen et al (unpublished data) randomized spider-fearful individuals to conditions of videotaped exposure to a spider in a specific location of a house or videotaped exposure to locations of a house without the spider. Measures included electrodermal responding and disgust ratings, because disgust is elevated in persons phobic of

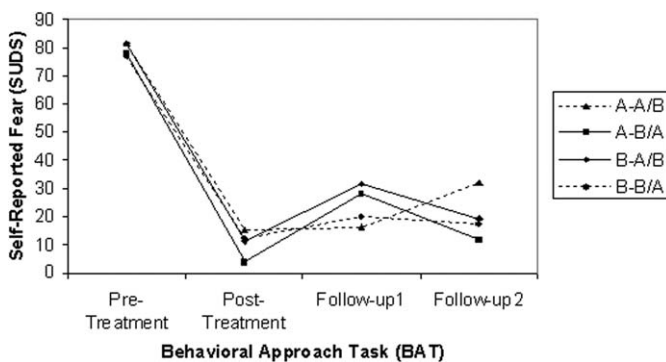


Figure 3. Mean self-reported fear (SUDS) for spider-fearful subjects treated in context A and tested in contexts A and B and subjects treated in context B and tested in contexts B and A. Follow-ups 1 and 2 refer to follow-up tests later that day, the first of which was in the same context as extinction treatment, while the second was in a different context. (Reproduced with permission from Mystkowski et al 2002). SUDS, subjective units of distress.

spiders (e.g., Matchett and Davey 1991; Tolin et al 1997). Not only did videotaped exposure to the spider lead to decreased electrodermal responding and disgust ratings compared with the control group, but both indices subsequently increased when tested with a videotape showing the same spider in a new context (e.g., a location in the house that was not previously seen) relative to the original exposure context.

Bouton and Swartzentruber (1991) suggested that mismatch of internal states during treatment and follow-up, manipulated via drug state, can also lead to significant return of fear. This has direct relevance to the use of psychotropic medications in combination with exposure therapy for anxiety disorders and may explain the high rates of relapse when medications are withdrawn (e.g., Marks et al 1993), particularly those medications that have a very distinct effect on internal state (i.e., high-potency benzodiazepines).

Initial failure to induce context-based renewal via alprazolam (Zoellner and Craske, unpublished data) was most likely attributable to dosage levels (.25 mg) that were too low to generate a sufficiently salient context. Evidence for internal context specificity of extinction was demonstrated in a study that manipulated drug state through caffeine versus placebo ingestion for individuals fearful of spiders (Mystkowski et al 2003). Participants received one session of exposure therapy under the influence of a randomly chosen drug condition (i.e., placebo or caffeine, 4 mg/kg of body weight) and were reassessed 1 week later after ingesting a drink mixture that was either the same or different than the drink ingested during the previous treatment session. As predicted, participants experiencing incongruent drug states exhibited significantly greater self-reported return of fear, measured during a behavioral approach task, from posttreatment to follow-up than those participants experiencing congruent drug states. Interestingly, the effects were comparable whether the shift was from caffeine to placebo or vice versa. However, the effect sizes (calculated as the difference between mean pretreatment and posttreatment scores divided by the pooled standard deviations) (Cohen 1988) were smaller than those achieved via the inside/outside context shifts (Mystkowski et al 2002; Mystkowski and Mineka, *in press*).

In summary, a series of experimental studies have consistently demonstrated context specificity effects of exposure therapy in circumscribed phobias, with larger effects obtained when more distinctly different contexts were employed. Conceivably, even larger effects might have been obtained if retesting had occurred in the original fear acquisition context. This is because in the animal literature the contextual control of return of fear is stronger when testing for renewal of fear occurs in the context in which the fear was originally conditioned than when tested in a new context (Bouton and Brooks 1993).

Attenuation of Context-Based Return of Fear

An obvious implication of these studies on the effects of contextual change on enhancing return of fear would seem to be to conduct exposure therapy in multiple contexts so that extinction memories can be cued by multiple contexts. Unfortunately, however, animal studies on this topic have revealed very inconsistent results, with only some finding that extinction in multiple contexts has reduced fear renewal in a final novel context relative to conducting extinction in only one context (see Bouton et al, *in press*, for a review). In a recent study of human fearful participants, Vansteenwegen et al (unpublished data) compared multiple context exposure treatment, in which spider-fearful individuals viewed video fragments of a spider in different locations of

a house, to single context exposure in which participants viewed fragments always in the same location. Although disgust ratings evinced context renewal effects in both groups on subsequent retesting to video fragments of a spider in a novel house location, electrodermal responding showed a renewal effect in the single context exposure group only.

Because it is not always feasible to conduct exposures in original fear acquisition contexts or multiple contexts, *Mystkowski et al (in press)* sought to investigate whether a contextual-based return of fear could be counteracted via “mental rehearsal.” Mental rehearsal of extinction contexts (e.g., the therapist, treatment information, and the physical surroundings where treatment took place) may override veridical context shifts. Such results might be anticipated based on findings in the verbal memory literature. Specifically, participants who are instructed to recall the original learning environment just prior to free recall of a list of words in an unfamiliar environment experience a release from contextual dependence otherwise observed following contextual shifts and perform identically to participants tested in the original learning environment (*Smith 1979*).

Spider-phobic participants were treated and followed up in the same context (i.e., matched context groups) or treated and followed up in different contexts (i.e., mismatched context groups). Half of the participants in each of these two groups were instructed to mentally rehearse the treatment context and the material learned in that context before entering the test context at follow-up, whereas the other half was instead asked to recall a neutral scenario. Self-report data replicated previous research on contextually driven return of fear, with strong effect sizes between groups and a high degree of statistical power. Furthermore, participants who mentally rehearsed the treatment context before encountering the phobic stimulus in a new context at follow-up had less return of fear than those who did not (see *Figure 4*). Indeed, their fear levels were comparable to those obtained in participants tested at follow-up in the original treatment context.

Other research has demonstrated the value of objects as retrieval cues in relation to reactivity to alcohol cues. Specifically, *Collins and Brandon (2002)* showed that the return of alcohol cue reactivity after extinction due to a context change could be

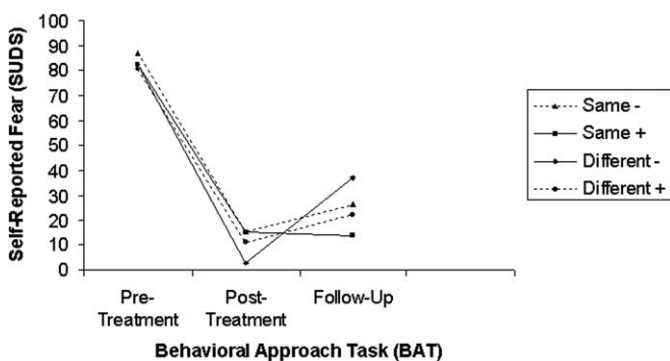


Figure 4. Mean self-reported fear (SUDS) for spider-fearful subjects from pretreatment to follow-up BATs. Group “Same” had matched treatment and follow-up contexts, while group “Different” had mismatched treatment and follow-up contexts. A plus sign (+) after the group name indicates treatment context mental reinstatement instructions, compared with a minus sign (–) for nontreatment context mental reinstatement instructions. (Reproduced with permission from *Mystkowski et al, in press*). SUDS, subjective units of distress; BAT, Behavioral Approach Test.

reduced by the use of retrieval cues that were present during extinction. However, mental rehearsal effects would seem to have more practical value than retrieval objects, given that fear cues might be encountered at times when retrieval objects are not available (*Mystkowski and Mineka, in press*). Notably, these human studies corroborate evidence from rodent studies showing that retrieval cues during extinction trials attenuate context-based renewal effects and more so than relatively novel cues or cues that were present during conditioning rather than extinction (*Brooks and Bouton 1994*).

Reinstatement Effects

The experimental investigation of reinstatement effects is prohibited by ethical constraints with human phobic samples. However, naturalistic observations support the notion of fear reinstatement. For example, *Wade et al (1993)* found that panic disorder and agoraphobia clients who experienced more negative life events 3 to 5 years after a 12-week behavioral treatment were functioning less well than those who experienced fewer negative life events. Similarly, relapsers from treatment for obsessive-compulsive disorder reported more distress over intervening life events since the end of treatment than clients who maintained their treatment gains (*Steketee 1993*). Notably, these observations suggest that reinstatement effects may occur with aversive events that are distinctly different from the aversive event involved in initial fear acquisition, as was found by *Rescorla and Heth (1975)* (Experiment 3) in rodents.

Discussion

Although most extinction research is conducted in animal laboratories, the study of extinction learning in human fear conditioning has received more attention during the last decade. We have reviewed experimental investigations of the impact of conditioned inhibitors, conditioned exciters, context renewal, and reinstatement on fear extinction in human samples. The results of these studies have generally confirmed predictions derived from animal research, with one exception being the effect of conditioned exciters during extinction training. First and foremost, however, the findings from human samples support the view that extinction does not entail unlearning and that the addition of conditioned inhibitors during extinction training, context changes, or the presentation of US-only trials after extinction training can lead to a partial recovery of the apparently “extinguished” fear responses. These findings raise the immediate question of how the procedure of extinction can be altered to improve its behavioral outcome. The work on extinction in multiple contexts and use of reminder cues has been important in this respect, not only because of its theoretical relevance but also because of the potential clinical applications.

Given that research has established comparable findings in human and nonhuman animals, we see a number of directions for future research. First, attention should be focused on those areas of divergence between research with humans and laboratory rodents. We already pointed to discrepancies concerning the effects of adding extra exciters during extinction (*Lovibond et al 2000; Rescorla 2000; Vervliet et al 2005*). Another example is AAB-renewal. Although the renewal effect is generally less pronounced in AAB designs (as compared with ABA designs) in animal studies, human extinction studies have failed to replicate the AAB effect (*Vansteenwegen et al 2005; Vervliet et al 2004*).

Irrespective of whether these discrepancies are replicable or not, a second important research focus will be to study corre-

spondences and divergences at the process level. Similar outcomes do not necessarily reflect similar processes. As already indicated, the role of higher-order, propositional processes in human conditioning deserves further exploration. Although no one will dispute that reasoning processes are an essential component of how humans approach the conditioning preparations in which they participate, the extent to which these propositions have a causal impact needs further examination. It is our view that these studies should not assume an all-or-none approach but should focus on the conditions under which propositional mechanisms have a key influence compared with the conditions under which their impact is less pronounced (Hermans and Van Gucht, in press).

The research that we reviewed stemmed from two related approaches: experimental laboratory research in nonclinical groups and experimental clinical work in anxious individuals. We believe that the combination of both approaches, together with animal conditioning work, provides a solid basis to answer the many research questions that we summarized in this discussion.

Preparation of this manuscript was (in part) based on research work supported by K.U. Leuven grant GOA/2001/01 "Extinction and the return of conditioned responses" (awarded to Drs. Eelen, Van den Bergh, and Baeyens), as well as by National Institute of Mental Health (NIMH) RFA-MH-04-005 "Developing translational research on mechanisms of fear extinction" (awarded to Drs. Barad and MGC) and NIMH R01 MH65651 "Common and specific risk factors for emotional disorders" (awarded to MGC and SM). In addition, preparation of this manuscript was supported by Australian Research Council grant A10007156 (to PFL).

Aspects of this work were presented at the conference, "Extinction: The Neural Mechanisms of Behavior Change," February 2–6, 2005, in Ponce, Puerto Rico. The conference was sponsored by the National Institute of Mental Health, National Institute of Drug Abuse, Ponce School of Medicine, University of Puerto Rico COBRE Program, Pfizer Global Pharmaceutical, and the Municipality of Ponce.

- Barad M (in press): The brain basis of fear extinction. In: Craske MG, Hermans D, Vansteenwegen D, editors. *Fear and Learning: From Basic Processes to Clinical Implications*. Washington, DC: American Psychological Association.
- Barlow DH, Craske MG (1994): *Mastery of Your Anxiety and Panic*. Albany, NY: Graywind Publications.
- Beaver JD, Mogg K, Bradley BP (2005): Emotional conditioning to masked stimuli and modulation of visuospatial attention. *Emotion* 5:67–79.
- Bouton ME (1988): Context and ambiguity in the extinction of emotional learning: Implications for exposure therapy. *Behav Res Ther* 26:137–149.
- Bouton ME (1994): Context, ambiguity, and classical conditioning. *Curr Dir Psychol Sci* 3:49–53.
- Bouton ME (2004): Context and behavioral process in extinction. *Learn Mem* 11:485–494.
- Bouton ME, Brooks DC (1993): Time and context effects on performance in a Pavlovian discrimination reversal. *J Exp Psychol Anim Behav Process* 19:165–179.
- Bouton ME, Mineka S, Barlow DH (2001): A modern learning theory perspective on the etiology of panic disorder. *Psychol Rev* 108:4–32.
- Bouton ME, Ricker ST (1994): Renewal of extinguished responding in a second context. *Anim Learn Behav* 22:317–324.
- Bouton ME, Swartzentruber D (1986): Analysis of the associative and occasion-setting properties of contexts participating in a Pavlovian discrimination. *J Exp Psychol Anim Behav Process* 12:333–350.
- Bouton ME, Swartzentruber D (1991): Sources of relapse after extinction in Pavlovian and instrumental learning. Special issue: Applied learning theory—Research issues for the 1990s. *Clin Psychol Rev* 11:123–140.
- Bouton ME, Woods AM, Moody EW, Sunsay C, Garcia-Gutierrez A (in press): Counteracting the context-dependence of extinction: Relapse and some tests of possible methods of relapse prevention. In: Craske MG, Hermans D, Vansteenwegen D, editors. *Fear and Learning: From Basic Processes to Clinical Implications*. Washington, DC: APA Books.
- Brooks DC, Bouton ME (1994): A retrieval cue for extinction attenuates response recovery (renewal) caused by a return to the conditioning context. *J Exp Psychol Anim Behav Process* 20:366–379.
- Carter RM, Hofstötter C, Tsuchiya N, Koch C (2003): Working memory and fear conditioning. *Proc Natl Acad Sci USA* 100:1399–1404.
- Clark RE, Squire LR (1998): Classical conditioning and brain systems: The role of awareness. *Science* 280:77–81.
- Cohen J (1988): *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.
- Collins BN, Brandon TH (2002): Effects of extinction context and retrieval cues on alcohol cue reactivity among nonalcoholic drinkers. *J Consult Clin Psychol* 70:390–397.
- Davey GCL (1997): A conditioning model of phobias. In: Davey GCL, editor. *Phobias: A Handbook of Theory, Research, and Treatment*. Chichester, England: Wiley.
- Dawson ME, Schell AM, Beers JR, Kelly A (1982): Allocation of cognitive processing capacity during human autonomic classical conditioning. *J Exp Psychol* 111:273–294.
- 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. *Psychol Bull* 127:853–869.
- De Houwer J, Vandorpe S, Beckers T (2005): On the role of controlled cognitive processes in human associative learning. In: Wills A, editor. *New Directions in Human Associative Learning*. Mahwah, NJ: Lawrence Erlbaum, 41–63.
- Delamater AR (1996): Effects of several extinction treatments upon the integrity of Pavlovian stimulus–outcome associations. *Anim Learn Behav* 24:437–449.
- Delamater AR (2004): Experimental extinction in Pavlovian conditioning: Behavioural and neuroscience perspectives. *Q J Exp Psychol B* 57B:97–132.
- Dirikx T, Hermans D, Vansteenwegen D, Baeyens F, Eelen P (2004): Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learn Mem* 11:549–555.
- Eelen P, Hermans D, Baeyens F (2001): Learning perspectives on anxiety disorders. In: Griez E, Faravelli C, Nutt D, Zohar J, editors. *Anxiety Disorders: An Introduction to Clinical Management and Research*. London: John Wiley and Sons, 249–264.
- Falls WA (1998): Extinction: A review of theory and evidence suggesting that memories are not erased with nonreinforcement. In: O'Donohue W, editor. *Learning and Behavior Therapy*. Boston: Allyn & Bacon, 205–229.
- Hermans D, Crombez G, Vansteenwegen D, Baeyens F, Eelen P (2002): Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In: Shohov SP, editor. *Advances in Psychology Research*, vol. 12. New York: Nova Science Publishers, 17–41.
- Hermans D, Dirikx T, Vansteenwegen D, Baeyens F, Van den Bergh O, Eelen P (2005): Reinstatement of fear responses in human aversive conditioning. *Behav Res Ther* 43:533–551.
- Hermans D, Van Gucht D (in press): Addiction: Integrating learning perspectives and implicit cognition. In: Wiers RW, Stacy AW, editors. *The Handbook of Implicit Cognition and Addiction*. Thousand Oaks, CA: Sage Publishers.
- Hermans D, Vansteenwegen D, Crombez G, Baeyens F, Eelen P (2002): Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behav Res Ther* 40:217–234.
- Knight DC, Smith CN, Cheng DT, Stein EA, Helmstetter FJ (2004): Amygdala and hippocampal activity during acquisition and extinction of human fear learning. *Cogn Affect Behav Neurosci* 4:317–325.
- Koster EHW, Crombez G, Van Damme S, Verschuere B, De Houwer J (2004): Does imminent threat capture and hold attention? *Emotion* 4:312.
- LaBar KS, Gatenby JC, Gore JC, LeDoux JE, Phelps EA (1998): Human amygdala activation during conditioned fear acquisition and extinction: A mixed-trial fMRI study. *Neuron* 20:937–945.

- LaBar KS, Phelps EA (2005): Reinstatement of conditioned fear in humans is context dependent and impaired in amnesia. *Behav Neurosci* 119:677–686.
- Lang PJ, Bradley MM, Cuthbert BN (1990): Emotion, attention, and the startle reflex. *Psychol Rev* 9:377–395.
- Lipp OV, Purkis HM (2005): No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cogn Emotion* 19: 269–282.
- Lipp OV, Siddle DAT, Dall PJ (1993): Effects of miscuing on Pavlovian conditioned responding and on probe reaction time. *Aust J Psychol* 45:161–167.
- Lovibond PF (2003): Causal beliefs and conditioned responses: Retrospective reevaluation induced by experience and by instruction. *J Exp Psychol Learn Mem Cogn* 29:97–106.
- Lovibond PF (2004): Cognitive processes in extinction. *Learn Mem* 11:495–500.
- Lovibond PF, Davis NR, O'Flaherty AS (2000): Protection from extinction in human fear conditioning. *Behav Res Ther* 38:967–983.
- Lovibond PF, Shanks DR (2002): The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *J Exp Psychol Anim Behav Process* 28:3–26.
- Marks IM, Swinson RP, Basoglu M, Kuch K, Noshirvani H, O'Sullivan G, et al (1993): Alprazolam and exposure alone and combined in panic disorder with agoraphobia: A controlled study in London and Toronto. *Br J Psychiatry* 162:776–787.
- Matchett G, Davey GCL (1991): A test of a disease-avoidance model of animal phobias. *Behav Res Ther* 29:91–94.
- Milad MR, Orr SP, Pitman RK, Rauch SL (2005): Context modulation of memory for fear extinction in humans. *Psychophysiology* 42(4):456–464.
- Mineka S (1985): Animal models of anxiety-based disorders: Their usefulness and limitations. In: Maser J, Tuma A, editors. *Anxiety and the Anxiety Disorders*. Hillsdale, NJ: Lawrence Erlbaum, 199–244.
- Mineka S, Mystkowski J, Hladek D, Rodriguez B (1999): The effects of changing contexts on return of fear following exposure treatment for spider fear. *J Consult Clin Psychol* 67:599–604.
- Mitchell CJ, Lovibond PF (2002): Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *Q J Exp Psychol B* 55B:311–329.
- Myers KM, Davis M (2002): Behavioral and neural analysis of extinction. *Neuron* 36:567–584.
- Mystkowski JL, Craske MG, Echiverri AM (2002): Treatment context and return of fear in spider phobia. *Behav Ther* 33:399–416.
- Mystkowski JL, Echiverri AM, Labus JS, Craske MG (in press): Mental reinstatement of context and return of fear in spider phobia. *Behav Ther*.
- Mystkowski JL, Mineka S (in press): Behavior therapy for fears and phobias: Context specificity of fear extinction. In: Baker TB, Bootzin R, Treat T, editors. *Psychological Clinical Science: Recent Advances in Theory and Practice: Integrative Perspectives in Honor of Richard M. McFall*. Hillsdale, NJ: Lawrence Erlbaum.
- Mystkowski JL, Mineka S, Vernon LL, Zinbarg RE (2003): Changes in caffeine state enhance return of fear in spider phobia. *J Consult Clin Psychol* 71:243–250.
- Öhman A, Mineka S (2001): Fears, phobias and preparedness: Toward an evolved module of fear and fear learning. *Psychol Rev* 108:483–522.
- Pavlov IP (1927): *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. London: Oxford University Press.
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004): Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron* 43:897–905.
- Phelps EA, O'Connor KJ, Gatenby JC, Gore JC, Grillon C, Davis M (2001): Activation of the left amygdala to a cognitive representation of fear. *Nat Neurosci* 4:437–441.
- Powers MB, Smits JAJ, Telch MJ (2004): Disentangling the effects of safety-behavior utilization and safety-behavior availability during exposure-based treatments: A placebo-controlled trial. *J Consult Clin Psychol* 72: 448–454.
- Rauch SL, Shin LM, Phelps EA (in press): Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research past, present and future. *Biol Psychiatry*.
- Rescorla RA (1996): Preservation of Pavlovian associations through extinction. *Q J Exp Psychol B* 49B:245–258.
- Rescorla RA (2000): Extinction can be enhanced by a concurrent excitator. *J Exp Psychol Anim Behav Process* 26:251–260.
- Rescorla RA (2001): Experimental extinction. In: Mowrer RR, Klein SB, editors. *Handbook of Contemporary Learning Theories*. Mahwah, NJ: Lawrence Erlbaum Associates, 119–154.
- Rescorla RA (2003): Protection from extinction. *Learn Behav* 31:124–132.
- Rescorla RA (2004): Spontaneous recovery. *Learn Mem* 11:501–509.
- Rescorla RA, Heth CD (1975): Reinstatement of fear to an extinguished conditioned stimulus. *J Exp Psychol Anim Behav Process* 104:88–96.
- Rescorla RA, Wagner AR (1972): A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors. *Classical Conditioning II: Current Research and Theory*. New York: Appleton-Century-Crofts 64–99.
- Ressler KJ, Rothbaum BO, Tannenbaum L, Anderson P, Graap K, Zimand E, et al (2004): Cognitive enhancers as adjuncts to psychotherapy: Use of D-cycloserine in phobias to facilitate extinction of fear. *Arch Gen Psychiatry* 61:1136–1144.
- Richardson R, Duffield TQ, Bailey GK, Westbrook RF (1999): Reinstatement of fear to an extinguished conditioned context. *Anim Learn Behav* 27:399–415.
- Rodriguez BI, Craske MG, Mineka S, Hladek D (1999): Context-specificity of relapse: Effects of therapist and environmental context on return of fear. *Behav Res Ther* 37:845–862.
- Salkovskis PM (1991): The importance of behaviour in the maintenance of anxiety and panic: A cognitive account. *Behav Psychother* 19:6–19.
- Sloan T, Telch MJ (2002): The effects of safety-seeking behavior and guided threat reappraisal on fear reduction during exposure: An experimental investigation. *Behav Res Ther* 40(3):235–251.
- Smith SM (1979): Remembering in and out of context. *J Exp Psychol Hum Learn Mem* 5:460–471.
- Sołtysik SS, Wolfe GE, Nicholas T, Wilson WJ, Garcia-Sanchez JL (1983): Blocking of inhibitory conditioning within a serial conditioned stimulus-conditioned inhibitor compound: Maintenance of acquired behavior without an unconditioned stimulus. *Learn Motiv* 14:1–29.
- Steketee G (1993): Social support and treatment outcome of obsessive compulsive disorder at 9-month follow-up. *Behav Psychother* 21: 81–95.
- Thomas BL, Ayres JJB (2004): Use of the ABA fear renewal paradigm to assess the effects of extinction with co-present fear inhibitors or excitators: Implications for theories of extinction and for treating human fears and phobias. *Learn Motiv* 35:22–52.
- Tolin DF, Lohr JM, Sawchuk CN, Lee TC (1997): Disgust and disgust sensitivity in blood-injection-injury and spider phobia. *Behav Res Ther* 35:949–954.
- Vansteenwegen D, Dirikx T, Hermans D, Vervliet B, Eelen P (in press): Renewal and reinstatement in human fear conditioning. In: Craske MG, Hermans D, Vansteenwegen D, editors. *Fear and Learning: From Basic Processes to Clinical Implications*. Washington, DC: American Psychological Association.
- Vansteenwegen D, Hermans D, Vervliet B, Francken G, Beckers T, Baeyens F, et al (2005): Return of fear in a human differential conditioning paradigm caused by a return to the original acquisition context. *Behav Res Ther* 43:323–336.
- Vervliet B, Vansteenwegen D, Baeyens F, Hermans D, Eelen P (2005): Return of fear in a human differential conditioning paradigm caused by a stimulus change after extinction. *Behav Res Ther* 43:357–371.
- Vervliet B, Vansteenwegen D, Eelen P (2004): Generalization of extinguished skin conductance responding in human fear conditioning. *Learn Mem* 11:555–558.
- Wade S, Monroe S, Michelson L (1993): Chronic life stress and treatment outcome in agoraphobia with panic attacks. *Am J Psychiatry* 10:1491–1495.
- Wells A, Clark DM, Salkovskis PM, Ludgate J, Hackmann A, Gelder M (1995): Social phobia: The role of in-situation safety behaviours in maintaining anxiety and negative beliefs. *Behav Ther* 26:153–161.
- Westbrook RF, Iordanova M, McNally G, Richardson R, Harris JA (2002): Reinstatement of fear to an extinguished conditioned stimulus: Two roles for context. *J Exp Psychol Anim Behav Process* 28:97–110.