Biological Psychology 92 (2013) 51–58

Contents lists available at SciVerse ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Extinction, generalization, and return of fear: A critical review of renewal research in humans

Bram Vervliet*, Frank Baeyens, Omer Van den Bergh, Dirk Hermans

Department of Psychology, University of Leuven, Belgium

ARTICLE INFO

Article history:
Received 18 April 2011
Accepted 12 January 2012
Available online 26 January 2012

Keywords:
Fear
Conditioning
Human
Renewal

ABSTRACT

The main behavioral signature of fear extinction is its fragility. This is exemplified by the renewal effect, where a change in the background context produces recovery of fear to a conditioned-and-extinguished stimulus. Renewal is the backbone of a widely accepted theory of extinction in animal research, as well as an important experimental model to screen novel treatment techniques. This has led to an explosion of fear renewal research in humans. However, the mere observation of return of fear in a renewal procedure is not sufficient to validate this particular theory of extinction in the tested sample/procedure. Here, we systematically outline a set of experimental tests that aid in evaluating alternative extinction/renewal mechanisms. We examine published renewal studies in human fear conditioning and conclude that the prevailing theory of extinction is often taken for granted, but critical tests are lacking. Including these tests in future research will not only reveal the fear extinction mechanism in humans, but also inspire further developments in extinction treatment research.

1. Introduction

Human conditioning research is strongly inspired by animal conditioning findings and theories. One of the central aims is to bridge the translational gap between pre-clinical animal research and patient studies, by testing the animal-based empirical findings and theoretical models in a human (healthy and/or patient) sample using comparable experimental methodology. Validating animal-based theories in human experimental protocols provides the necessary ground for generalizing innovative applications from animals to patients.

Fear extinction is a major research focus in animal conditioning, and has culminated in a widely used theory of extinction (Bouton, 1994). It is obviously also very important in the clinical domain. Fears constitute an important part of anxiety disorders; the extinction of fears contributes importantly to treatment success (Craske et al., 2008). Currently, Bouton's theory of extinction is also widely applied in human clinical and pre-clinical research. It serves as an important guideline for investigating fear extinction on different levels (e.g., emotion, cognition, brain imaging) and for developing and screening extinction-enhancing techniques (see Vervliet, 2008). Most of these studies index the effects of extinction training by assessing contextual renewal, which is also the central phenomenon on which Bouton's theory of extinction is based. In this paper, we (1) describe the renewal phenomenon and outline Bouton's theory of extinction, (2) summarize the critical tests that have validated this extinction theory in animal conditioning research, and (3) evaluate its empirical basis in human fear conditioning. It will become clear throughout the paper that, theoretically, different mechanisms of extinction may (co-)exist and lead up to renewal, but that critical tests are currently lacking in the human domain. Hence, detailed behavioral analysis of extinction in the human fear conditioning paradigm is necessary to further develop human theories of fear extinction, to guide interpretations of brain activity patterns underlying extinction, and to design and screen new behavioral and/or pharmacological enhancers of extinction.

1.1. Fear conditioning and extinction

In general, the term 'fear conditioning' refers to an experimental procedure in which (1) a neutral stimulus is arranged as a reliable predictor of an aversive stimulus (mostly an electrical stimulation), and (2) the predicting stimulus starts eliciting cognitive, emotional and/or behavioral reactions in anticipation of the aversive shock. In contemporary research, psychophysiological recordings and subjective ratings are often combined in order to get a clear picture of the conditioned fear response. These recordings frequently include skin conductance responses and/or fear-potentiated startle reflexes. Ratings often ask for the expectancy of the shock, the level

* Corresponding author at: Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium.
E-mail address: bram.vervliet@psy.kuleuven.be (B. Vervliet).

0301-0511/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.biopsycho.2012.01.006

of fear, or the acquired valence of the predicting stimulus (see Lipp, 2006).

The conditioning process is usually conceptualized as follows. Paired presentations of a neutral stimulus (the conditional stimulus, CS) and an aversive stimulus (the unconditional stimulus, US) result in the formation of an association between the memory representations of the neutral CS and the aversive US. Future confrontations with the CS will activate its own representation and, by virtue of the association, the memory of the US as well. This 'thinking of the aversive US' elicits anticipatory fear. Hence, the basic associative learning framework is a sort of 'spreading of activation' framework.

Fear extinction occurs when repeated presentations of the CS, in the absence of the US, lead to a gradual decay of anticipatory fear reactions. This apparently simple phenomenon has proven difficult to explain. Since Pavlov's original observations (Pavlov, 1927), the crucial question has been whether behavioral extinction reflects *unlearning* of the original CS–US association (e.g., Rescorla and Wagner, 1972), or whether it reflects an *inhibition* of the original CS–US association (e.g., Pavlov, 1927; Konorski, 1967). In the former case, extinction is considered a passive process by which the CS returns to a neutral state (as if no conditioning has occurred). However, numerous studies demonstrate that fear reactions can easily return to a conditioned-and-extinguished CS (Bouton, 2002). This strongly suggests that an extinguished CS is not at all neutral: The fear reactions can only recover *if* the original CS–US association somehow survives extinction. Arguably, new learning occurs during extinction that inhibits, but not erases, the CS–US association (Pavlov, 1927; Konorski, 1967). This new learning is often conceptualized as the formation of an inhibitory CS–US association. Henceforth, the CS both activates and deactivates the representation of the US, so that the CS will sometimes elicit fear and sometimes not. The circumstances at test largely determine which association will determine the behavioral outcome of the CS (fear/no fear). This is convincingly demonstrated by the renewal effect.

2. Contextual renewal: extinction, generalization, and return of fear

Contextual renewal refers to a situation where changes in the background context evoke a recovery of fear to a conditioned-and-extinguished stimulus (see Bouton, 2002). For example, if CS–US pairings have occurred in context A, and CS-only extinction presentations have occurred in context B, renewal tests can consist of presenting the CS in context A again or in a novel context C (ABA-renewal and ABC-renewal, respectively; Bouton and Bolles, 1979). The typical renewal observation is an increase of fear reactions elicited by the CS. A third variant comprises CS–US pairings and CS-only extinction presentations in the same context A, followed by tests in a novel context B (AAB-renewal; Bouton and Ricker, 1994). Control conditions typically consist of AAA (all phases in the same context) and/or ABB procedures (final test in the extinction context). Although all three types of renewal have been documented in the literature, ABC-renewal is often smaller than ABA-renewal (e.g., Harris et al., 2000) and AAB-renewal is not always observed (e.g., Bouton and King, 1983). For excellent comprehensive reviews on renewal in animal fear conditioning research, readers are referred to papers by Bouton (e.g., 1994, 2002) for behavioral analysis, and to Myers and Davis (2002) and Quirk and Mueller (2008) for neural analysis.

The behavioral theory of extinction (Bouton, 1994) is strongly inspired by these contextual renewal phenomena. The observed sensitivity of extinction to contextual changes has led to the assumption that activating the second-learned, inhibitory CS–US association requires input from the CS and the extinction context *in concert*. Hence, presentations of the CS in any other context will lack the necessary context-input and therefore *fail to activate* the inhibitory association. This leaves the way open for the excitatory association to activate the US representation in memory, which then leads to a recovery of the conditioned reaction.

The bottom line of the model is that the context is merged in the representation of the extinction memory (CS–noUS), but not the acquisition memory (CS–US). This is in line with the observation that conditioned fears are generally context-independent, whereas extinction effects are generally context-specific. Importantly, the theory states that the extinction context does not directly activate or inactivate the US representation; it *modulates* the inhibitory CS–US association. From another perspective, the extinction context *retrieves* the specific CS extinction memory (the inhibitory CS–US association). This has important implications for the conditions under which extinction and renewal are predicted to occur. These conditions represent the critical tests of the theory.

3. The theory of extinction: critical tests

The extinction theory of Bouton (1994) has been validated in animal conditioning research by systematically rejecting a set of alternative hypotheses that are simpler in nature (and hence, more parsimonious). We briefly introduce these alternative hypotheses and present the critical tests.

3.1. Incomplete extinction

The hypothesis. A fear reaction to a conditioned-and-extinguished stimulus after a context change may simply be due to incomplete extinction.

The critical tests. First, a control group without context change between extinction and test provides a between-groups comparison at test. Second, the fear recovery has to be compared to the level of fear on the last extinction trial(s) in the context change group. Finally, these between- and within-group comparisons should produce a significant 2×2 interaction between group (ABA versus AAA) and moment (end of extinction versus test) (TEST 1).

3.2. Generalization decrement

The hypothesis. Under the assumption that a discrete stimulus (the CS) can be perceived differently in distinct contexts, stimulus generalization processes may underlie the renewal effect (see Lovibond et al., 1984). Consider ABA-renewal. The test occurs in the exact situation of acquisition (CS-in-A), but the extinction effect has to be generalized from CS-in-B to CS-in-A. It is well known that generalization decreases to progressively dissimilar stimuli (e.g., Lissek et al., 2008). Hence, the generalization of extinction from B to A is likely to be smaller than the original acquisition in A. The net result is an increase of fear evoked by the CS when presented in A. This is a very important explanation, because it is very parsimonious: it follows directly from the oldest principles in conditioning research.

The critical tests. If stimulus generalization is at the heart of the renewal phenomenon, an equal generalization decrement should emerge on the transition from A to B after acquisition. Bouton and colleagues routinely do not observe a fear decrement on the first extinction trial(s) in B (e.g., Bouton and King, 1983). Apparently, the context change (from A to B) leaves acquisition intact, whereas the opposite context change (from B to A) disrupts extinction. This asymmetry has been the crucial challenge in renewal research, as it does not follow from a simple generalization mechanism.

But what if one observes a relative asymmetry, with a decrement after acquisition that is smaller than the recovery at test? At first sight, this seems to prove that fear acquisition generalizes more