Psychological Bulletin
2010, Vol. 136, No. 3, 390–421

0033-2909/10/$12.00 DOI: 10.1037/a0018916

Evaluative Conditioning in Humans: A Meta-Analysis

Wilhelm Hofmann
University of Würzburg and University of Amsterdam

Jan De Houwer
Ghent University

Marco Perugini
University of Milan—Bicocca

Frank Baeyens
University of Leuven

Geert Crombez
Ghent University

This article presents a meta-analysis of research on *evaluative conditioning* (EC), defined as a change in the liking of a stimulus (conditioned stimulus; CS) that results from pairing that stimulus with other positive or negative stimuli (unconditioned stimulus; US). Across a total of 214 studies included in the main sample, the mean EC effect was $d = .52$, with a 95% confidence interval of .466–.582. As estimated from a random-effects model, about 70% of the variance in effect sizes were attributable to true systematic variation rather than sampling error. Moderator analyses were conducted to partially explain this variation, both as a function of concrete aspects of the procedural implementation and as a function of the abstract aspects of the relation between CS and US. Among a range of other findings, EC effects were stronger for high than for low contingency awareness, for supraliminal than for subliminal US presentation, for postacquisition than for postextinction effects, and for self-report than for implicit measures. These findings are discussed with regard to the procedural boundary conditions of EC and theoretical accounts about the mental processes underlying EC.

Keywords: evaluative conditioning, affective learning, attitude learning, associative learning, propositional learning

One of the most influential ideas in psychology is that human behavior is, to a large extent, governed by likes and dislikes (Allport, 1935; Martin & Levey, 1978). For instance, people prefer the company of people they like and try to avoid those they do not like; people buy and consume products they like rather than those they dislike; and they vote for and support politicians and ideas that they find sympathetic rather than repelling. Furthermore, preferences influence attention, memory, and judgments and form the basis of our emotional life (Fox, 2009). Given the pervasive impact of preferences on behavior, it is vital for our discipline to understand how preferences are formed and how they can be influenced. Although some likes and dislikes may be genetically determined (Poulton & Menzies, 2002), the vast majority of our preferences are learned rather than innate (Rozin & Millman, 1987). But precisely how humans acquire their likes and dislikes continues to be the subject of vigorous debate (Rozin, 1982; De Houwer, Thomas, & Baeyens, 2001).

The present article provides a meta-analysis of research on one possible manner in which likes and dislikes can be learned: *evaluative conditioning* (EC), which may be best defined as an effect that is attributed to a particular core procedure. Specifically, EC refers to a change in the valence of a stimulus (the effect) that is due to the pairing of that stimulus with another positive or negative stimulus (the procedure) (De Houwer, 2007a; De Houwer et al., 2001). The first stimulus is often referred to as the *conditioned stimulus* (CS), and the second stimulus is often referred to as the *unconditioned stimulus* (US). Typically, a CS becomes more positive when it has been paired with a positive US and more negative when it has been paired with a negative US. EC is a form of Pavlovian conditioning in that it involves a change in the responses to the CS that results from pairing the CS with a US. Whereas Pavlovian conditioning can refer to a change in any type of

Wilhelm Hofmann, Department of Psychology, University of Würzburg, Würzburg, Germany, and Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; Jan De Houwer and Geert Crombez, Department of Psychology, Ghent University, Ghent, Belgium; Marco Perugini, Department of Psychology, University of Milan—Bicocca, Milan, Italy; Frank Baeyens, Department of Psychology, University of Leuven, Leuven, Belgium.

The present research was supported by a grant from the European Association of Social Psychology, from the Universitätsstiftung of the University of Würzburg, and from the Psychology Department of the University of Koblenz-Landau to Wilhelm Hofmann, by Grant BOF/GOA2006/001 of Ghent University to Jan De Houwer and Geert Crombez, and by Grant GOA/2007/03 of the University of Leuven to Frank Baeyens.

We thank Katharina Fischer and Anne Schwab for their help in coding and Umut Özdemir, Charlotte Schwab, and Judith Phieler for their help in retrieving literature and organizing the database. We also thank Roland Deutsch, Bertram Gawronski, Katie Lancaster, Siegfried Sporer, and David Wilson for valuable comments on the work reported in this article.

Correspondence concerning this article should be addressed to Wilhelm Hofmann, Department of Psychology, University of Würzburg, Roentgenring 10, 97070 Würzburg, Germany. E-mail: hofmannw@psychologie.uni-wuerzburg.de

response, EC concerns only a change in the evaluative responses to the CS, that is, a change in the liking of the CS (see De Houwer, 2007a, for an in-depth discussion).

A Short History of Evaluative Conditioning Research and Debates

The first demonstrations of EC effects date back more than 50 years (Razran, 1954; C. K. Staats & Staats, 1957). C. K. Staats and Staats (1957), for instance, showed that nonsense words that were paired with either positive or negative words acquired the same affective value of the words with which they were paired (see Jaanus, Defares, & Zwaan, 1990, for a review). Modern EC research was sparked by the work of Levey and Martin (1975). These authors introduced the so-called picture–picture paradigm that is still frequently used today. Participants first sorted a set of postcard pictures into liked, disliked, and neutral categories. In a subsequent acquisition phase, initially neutral postcards (CS) were presented together with liked, disliked, or other neutral postcards (USs). Subsequent liking ratings showed that the valence of the CSs that were paired with a liked or disliked US had changed in the respective direction of the US valence.

Since these early demonstrations, EC has been examined in a large number of areas. These include learning psychology (e.g., Martin & Levey, 1978), social psychology (e.g., Olson & Fazio, 2001; Walther, 2002), consumer science (e.g., Allen & Janiszewski, 1989; Stuart, Shimp, & Engle, 1987), emotion research (e.g., Mallan & Lipp, 2007; Niedenthal, 1990), neuroscience (Coppens et al., 2006; Everhart & Demaree, 2003), nutrition research (e.g., conditioned taste aversion learning; Bernstein & Webster, 1980), and clinical psychology (e.g., fear conditioning; Hermans et al., 2004; Olatunji, Lohr, Sawchuk, & Westendorf, 2005; for reviews on EC, see De Houwer et al., 2001; Field, 2005; and De Houwer, in press).

From a general perspective, research on EC has been guided by three main questions: First, a majority of the studies examined whether EC is a genuine (in the sense of "true," "authentic," and "replicable") and general phenomenon. Second, researchers investigated whether EC is a unique form of Pavlovian conditioning. They did so by trying to identify variables that influence EC and Pavlovian conditioning in different ways. The third question concerned the processes that underlie EC. Although several theories about the nature of these processes have been put forward over the years, relatively little research has been directed toward distinguishing among these models in an empirical manner. In the following sections, we provide a brief review of the relevant literature for each of these questions.

Literature Review

Genuineness and Generality of Evaluative Conditioning

Although a proportion of the older evidence for the existence of EC is compromised by a lack of appropriate controls (see Field & Davey, 1999), more recent studies have confirmed that EC is a genuine phenomenon that is observed under a variety of conditions (e.g., De Houwer, Baeyens, & Field, 2005). At the same time, failures to observe EC have been haunting the field (e.g., Field & Davey, 1999; Rozin, Wrzesniewski, & Byrnes, 1998). These findings suggest that EC may be subject to boundary conditions that have yet to be identified.

Evaluative Conditioning and Pavlovian Conditioning

Whereas some researchers have argued that EC does substantially differ from other forms of Pavlovian conditioning (e.g., Baeyens & De Houwer, 1995; Martin & Levey, 1994), others have raised doubts about this claim (e.g., Davey, 1994b; Lipp & Purkis, 2005). The debate on the uniqueness of EC has mainly focused on the impact of contingency awareness and extinction procedures (i.e., CS-only trials after acquisition) on EC. Several studies have reported EC even when participants were not aware of the CS–US contingencies (e.g., Baeyens, Eelen, & Van den Bergh, 1990; Dickinson & Brown, 2007; Walther & Nagengast, 2006). Such findings are remarkable because other forms of Pavlovian conditioning are dependent on the awareness of the CS–US contingencies (see Lovibond & Shanks, 2002, and Mitchell, De Houwer, & Lovibond, 2009b, for reviews). Other studies, however, have indicated that EC also occurs only after the participants become aware of the contingency between the CS and the US with which it was paired (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl, Unkelbach, & Corneille, 2009).

With regard to the impact of extinction procedures, several studies have found that the magnitude of EC was unaffected by the presence of unpaired CS presentations that occurred after the CS–US pairings (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Díaz, Ruiz, & Baeyens, 2005). Others, however, have reported data showing that unpaired CS presentations do significantly reduce EC effects (e.g., Lipp, Oughton, & LeLievre, 2003). Additionally, uncertainty remains about the impact of the statistical contingency between the CS and the US during acquisition, that is, about the impact of unpaired CS and US presentations that are intermixed with CS–US pairings during the learning phase (e.g., Baeyens, Hermans, & Eelen, 1993). In sum, it is still unclear whether there are variables that have a different impact on EC than on other forms of Pavlovian conditioning.

Theoretical Accounts of Evaluative Conditioning Effects

Little progress has been made in answering the third question: What is the nature of the mental processes responsible for EC? Several theoretical accounts of EC have been proposed. Because an exhaustive treatment is beyond the scope of this meta-analysis (for reviews, see De Houwer et al., 2001; De Houwer, in press), we briefly sketch five main accounts of EC (see Table 1 for an overview). In doing so, we point out some predictions that can be derived from each account.

The referential account. Baeyens, Eelen, Crombez, and Van den Bergh (1992) postulated that there are two types of Pavlovian conditioning, both of which are based on simple mechanisms of association formation in memory. The first type concerns the associative learning of predictive relations by which the CS becomes a signal for the upcoming presentation of the US. This type of signal or expectancy learning is hypothesized to be determined by the statistical contingency between the CS and the US. It is assumed to underlie most cases of Pavlovian conditioning (see also

Table 1
Overview of Theoretical Accounts of Evaluative Conditioning (EC)

Account	Type of process	Main assumptions	Primary predictions
Referential account (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992)	Automatic formation of associations between CS and US representation	● Distinguishes two types of learning: (1) Learning of predictions (through statistical contingency) by which a CS becomes a signal for the US (as in Pavlovian conditioning) (2) Learning of referential relations (through stimulus co-occurrence) by which CS becomes a stimulus that simply refers to the US without becoming a signal for the US. This type of learning is insensitive to statistical contingency ● EC is assumed to depend on the second type of learning	● EC increases with number of CS–US co-occurrences ● EC independent of contingency awareness ● EC resistant to extinction
Holistic account (Levey & Martin, 1975; Martin & Levey, 1978, 1994)	Automatic formation of holistic CS–US representation	● Co-occurrence leads to the enduring formation of a holistic representation that encodes stimulus features of both the CS and US, as well as the valence of the US ● CS can associatively activate this representation and thus the evaluation associated with the US	● EC increases with number of CS–US co-occurrences ● EC independent of contingency awareness ● EC resistant to extinction
Implicit misattribution account (Jones, Fazio, & Olson, 2009)	Automatic formation of associations between CS representation and evaluative response evoked by US	● Transfer of valence is the result of an automatic (in the sense of unintentional and unconscious) misattribution process such that the evaluative reaction evoked by the US becomes associated with CS. Hence, participants (incorrectly) assume that evaluation they experience is caused by the CS and not by the US	● Any variable that influences the likelihood of misattribution should influence EC (e.g., perceptual similarity; spatial proximity; mild rather than strong US valence) ● Contingency awareness may counteract misattribution
Conceptual categorization account (Davey, 1994b; Field & Davey, 1999)	Recategorization of CS as the result of highlighting the similarity between CS and US	● Pairing makes salient those features that it has in common with a liked or disliked US ● This selective highlighting of certain stimulus features can lead to a change in the categorization of a formerly neutral stimulus as “liked” or “disliked,” respectively	● EC increases with number of CS–US co-occurrences ● EC resistant to extinction ● EC increases with degree of perceptual similarity (e.g., feature overlap) between CS and US
Propositional account (De Houwer, 2007a, 2009a; Mitchell, De Houwer, & Lovibond, 2009a)	Formation and truth evaluation of propositions about the CS–US relation	● EC results from the formation of propositions about the CS–US relation ● Liking changes only after individuals have formed a conscious proposition that a CS is paired with a valenced US ● Propositional knowledge about CS–US relation can function as a justification for determining liking of the CS	● Contingency awareness is a necessary precondition for EC

Rescorla, 1988). The second type concerns the associative learning of merely referential relations. In referential learning, the CS becomes a stimulus that simply activates a mental representation of the US, without creating an expectancy that the US will appear. This is similar to the way that, for instance, reading the name of a beloved one may make one think of a kiss without necessarily expecting a kiss to occur. Referential learning is assumed to be determined by the mere co-occurrence of stimuli, not statistical contingency. It is this type of learning that is supposed to underlie EC (Baeyens et al., 1992).

Because referential learning is driven by the co-occurrence of stimuli (rather than by the statistical contingency of events), the referential account predicts that EC is resistant to extinction, that is, impervious to the effects of CS-only trials that are presented after CS–US trials (i.e., after acquisition). Finally, because referential learning is assumed to be part of a primitive automatic association formation mechanism, the referential account postulates that explicit awareness of CS–US contingencies is not necessary for EC to occur.

The holistic account. According to the holistic account (Levey & Martin, 1975; Martin & Levey, 1978, 1994), the co-occurrence of a CS and a US automatically results in the formation of a holistic representation, which encodes stimulus elements of both the CS and the US, as well as the valence of the US. Once the holistic representation has been formed, the CS can activate this representation and thus the evaluation that was associated with the US. Although Martin and Levey (1994) did not use the term *associative*, the model can be considered as quasi-associative. This is because it is hypothesized that, similar to the process of pattern completion in distributed associative networks, the CS can automatically "activate the larger information structure of which it has become a component" (Martin & Levey, 1994, p. 304).

The holistic model predicts that conditioned changes in liking depend mainly on CS–US co-occurrences. Subsequent CS-only trials should not alter the holistic representation and thus also should not alter the conditioned change in liking. Therefore, EC is assumed to be resistant to extinction. Just like the referential account, the holistic account also predicts that EC does not depend on awareness of CS–US contingencies.

The implicit misattribution account. Recently, Jones, Fazio, and Olson (2009) proposed a misattribution theory that has much in common with the holistic account of Martin and Levey (1978). According to Jones et al., it is the evaluative reaction evoked by the US that becomes associated with the CS. The resulting associative representation can be seen as equivalent to a holistic representation that contains the stimulus features of the CS and only the evaluative response component of the US. Like Martin and Levey, Jones et al. postulated that EC can be formed in the absence of awareness of the CS–US relation. Jones et al. did specify, however, that EC depends on an "implicit misattribution" process: Because affective reactions to stimuli are elusive phenomenological experiences and because the actual source of such experiences often may not be clear (Russell, 2003), evaluative responses to the US are likely to become incorrectly attributed to the CS during conditioning. Such an implicit misattribution of affective experiences is assumed to occur at an early stage of perceptual–cognitive processing and therefore does not depend on the conscious, explicit evaluation of the CS or US. However, any variable that influences the likelihood that the US valence will be misattributed to the CS should also influence EC. Jones et al. indeed observed an impact of a number of these variables, including spatial proximity (i.e., feelings are more likely to be attributed to CSs that are close in space to a US). The misattribution account also predicts that increasing degrees of feature overlap between the US and the CS should render misattribution more likely (Jones et al., 2009). Hence, EC effects should be larger for stimuli matched in perceptual similarity and for stimuli of the same rather than a different modality. Moreover, mildly valenced USs should result in stronger EC effects than strongly valenced USs. This is because the feelings evoked by strongly valenced USs should be less susceptible to source confusion. Finally, implicit misattribution may work best in the case of low contingency awareness because "such awareness could make salient the US and its evaluative aspects, possibly discouraging misattributions to the CS" (Jones et al., 2009, p. 944). Hence, whereas the referential and the holistic account predict that EC is independent of contingency awareness, the misattribution account predicts a negative relationship between the two.

The conceptual categorization account. According to Davey (1994a; see also Field & Davey, 1999), EC may not be the result of the formation of associations in memory but rather a result of conceptual learning. Specifically, a change in the liking of the CS may occur because the pairing of the CS and the US makes salient those features of the CS that it has in common with the US. As a result, the CS is more likely to be categorized as a liked (or disliked) stimulus. For example, imagine an evaluatively neutral face that has the features of brown eyes, long shape, full lips, and long hair. Consider now that this neutral face is repeatedly presented together with a liked US that has the features of blue eyes, round shape, full lips, and long hair. According to this account, the CS–US pairings will increase the salience of the features that the CS has in common with the US (i.e., full lips and long hair). Because of this, the pairing may change the evaluation of the face to the extent that it will be categorized as a liked stimulus.

The model of Davey (1994a) predicts that EC should depend mainly on the number of co-occurrences of the CS and US (rather than their statistical contingency) because it is on these trials that the salience of the CS features can change. Once the salience of certain CS features has been increased, these changes in salience (and thus liking) might persist even when the CS or the US is subsequently presented on its own. Hence, EC should be resistant to extinction. Finally, EC effects should be restricted to cases in which the CS and the US have features in common. Hence, EC effects are not expected (or at least are less likely) when the CS and the US belong to different modalities.

The propositional account. The propositional account considers the possibility that all forms of associative learning, including EC, depend on the nonautomatic formation and truth evaluation of propositions about CS–US relations (De Houwer, 2007a, 2009b; De Houwer et al., 2005; see also Mitchell et al., 2009b). Propositions can be defined as statements about a state of affairs in the world that can differ in the degree to which they are believed to be accurate. Applied to EC, the propositional account holds that the liking of the CS will change only after participants have formed the conscious proposition that the CS is paired with (or co-occurs with) a valenced US. Although the model does not always explain how this propositional knowledge results in a change in liking (see Mitchell, De Houwer, & Lovibond, 2009a), it does postulate that the formation of a proposition about the

CS–US relation is a necessary mediating step. One possible way this could occur is that participants use propositional knowledge about the CS–US relation to determine how much they like the CS. For instance, the fact that a CS is paired with a negative US can be seen as a justification for disliking the CS (De Houwer et al., 2005).

Because the formation of propositions is assumed to be a higher order, conscious, and effortful mental process, the propositional account predicts that EC should depend on awareness of the CS–US contingencies (and all variables that promote or hinder the formation of contingency awareness). Furthermore, EC should be moderated by variables related to the capacity to form propositions (e.g., sufficient processing resources). This view is thus not able to account for EC in the absence of contingency awareness. Furthermore, the propositional account is relatively mute with regard to the role of statistical consistency in comparison with co-occurrence of stimuli. On the one hand, it is possible that EC is mediated by propositions about the statistical contingency between the CS and the US because higher degrees of contingency should strengthen the belief that the presentation of CS and US is related. On the other hand, it is also possible that the propositions that underlie EC are limited to the fact that the CS and the US co-occur.

Summary of the theoretical accounts. The theoretical accounts described above differ in their assumptions about the mental processes that underlie EC effects. Most of them (referential account, holistic account, implicit misattribution account) more or less explicitly focus on the formation of associations in memory between elements of the CS and US representation and on the conditions that affect such memory formations. Once this association has been formed, the CS can activate the liking that was originally evoked by the US, thus leading to a change in liking of the CS. In contrast, the conceptual categorization account and the propositional account emphasize the role of higher order mental processes in EC effects, such as conceptual categorization and formation of propositions, respectively. Even though the five models differ in their assumptions, emphasis, and scope, it appears that none of them is yet formalized and sophisticated enough as to offer a comprehensive account of EC. This is also the reason why deriving specific predictions about how a given variable may moderate EC effects is often a difficult enterprise and necessitates the introduction of auxiliary assumptions. Nevertheless, there are some diverging predictions about key variables (see Table 1 for an overview). Perhaps most centrally, these concern (a) whether contingency awareness is independent from, facilitates, or even impedes EC; (b) whether EC is resistant to extinction; (c) the degree to which EC is influenced by the statistical contingency as opposed to the number of CS–US co-occurrences; and (d) the role of perceptual similarity or feature overlap more generally.

Conclusion of the Literature Review

Taken together, despite the importance of EC and the large number of studies that have examined this phenomenon, little is known about the generality of EC, its distinctiveness from other forms of learning, and the nature of the theoretical mechanisms underlying it. In large part, this state of affairs is due to the many conflicting findings that have been reported in the literature. Perhaps then, a quantitative synthesis of EC research is needed.

The Present Meta-Analysis

As in other fields of psychology, a large degree of ambiguity among primary empirical studies in a given domain suggests that effect sizes are not homogeneous but rather are dependent on certain boundary conditions or moderator variables. A tool that has proved useful in the social sciences in structuring the findings from primary research is the method of meta-analysis. Meta-analyses provide quantitative summaries of the available evidence in a field. They thus may offer a better basis for resolving debates that have a high level of empirical ambiguity and for directing future research (e.g., Cooper & Hedges, 1994; Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Lipsey & Wilson, 2001). To use a metaphor, meta-analysis is akin to what wanderers often do to better orient themselves: climb a hill for a better view of the landscape beneath. As such, we consider the method of meta-analysis as a useful way to gain a clearer picture of EC effects. Despite the large empirical catalog of primary studies that have been conducted on the subject, the field has never been summarized meta-analytically.

For the present meta-analysis we had three objectives: First, we wanted to provide an estimate for the overall magnitude of EC effects across a wide range of primary studies. Second, we wanted to assess the degree of heterogeneity in EC research, both in a descriptive manner by coding for a large set of procedural variations of EC, and in a quantitative manner by estimating the degree of heterogeneity in EC effects. Assessment of heterogeneity allows one to judge whether substantial variability due to moderator variables exists or whether all observed EC effects stem from one fixed population, varying across studies only as a result of sampling error (e.g., Hunter & Schmidt, 1990).

Third, given that substantial heterogeneity in effect sizes exists, the role of potential moderators can be investigated. In the present context, we were interested in whether EC effects vary as a function of a large set of procedural characteristics. The left column of Table 2 presents an overview of the potential moderators we identified. The list of moderators was adapted from an initial, unpublished meta-analytic pilot project conducted by Baeyens and Crombez (1994) and further developed for the present purposes. Another set of moderators, such as contingency awareness, subliminal stimulus presentation, and the role of extinction, was inferred from a range of conceptual review articles that have appeared in high-impact journals over the last decades (e.g., De Houwer et al., 2001, 2005; Field, 2000; Levey & Martin, 1990).

Rather than discussing the moderators in a more or less haphazard manner, we decided to organize them according to the heuristic framework for learning research that was recently provided by De Houwer (2009a, in press). This framework allows one to classify individual EC procedures on the bases of abstract aspects of the relation between stimuli and of concrete aspects of the way in which the relation is implemented. Abstract aspects encompass the statistical properties of the relation between the stimuli that are paired (e.g., the degree of statistical contingency) and possible changes in those properties (e.g., an acquisition phase of CS–US pairings followed by an extinction phase of unpaired CS presentations). Concrete aspects, in contrast, refer to the implementation of the relation between CS and US. The concrete implementation requires choices about (a) the organism that experiences the relation, (b) the properties of the stimuli that are

Table 2
Coded Study Characteristics and Interrater Agreement

Variable	Definition	Coding options	IA
	Concrete aspects (implementation of the relation)		
(a) Organism			
Sample	Sample under investigation	Normal participants (1–6): 1 = psychology student; 2 = other students; 3 = mixed (students and nonstudents); 4 = nonstudents; 5 = teenagers (ages 12–17); 6 = children (ages 6–11); 7 = pathological cases	.91
Gender	Proportion of female participants per group	Continuous	.95
(b) Stimulus properties			
CS modality	Of what modality was the CS?	1 = visual; 2 = auditory; 3 = taste/flavor; 4 = odor; 5 = verbal sensical; 6 = verbal nonsensical (e.g., nonsense syllables); 7 = haptic	.79
US modality	Of what modality was the US?	1 = visual; 2 = auditory; 3 = taste/flavor; 4 = odor; 5 = verbal sensical; 6 = verbal nonsensical; 7 = electrocutaneous stimulation; 8 = haptic	.90
CS–US modality match	Were CS and US of the same or of a different modality?	1 = yes (unimodal); 2 = no, CS and US were of a different modality (cross-modal)	—[a]
CS selection	How was CS selected regarding neutrality and individual assignment?	1= it was ensured that CS was neutral in valence and that assignment of CS was based on the participant's individual ratings; 2 = neutral valence pretest, not individually assigned; 3 = CS had initial valence, not individually assigned; 4 = CS had initial valence, individually assigned	.96
US selection	How was US selected regarding valence and individual assignment?	1 = US valenced was ensured via pretest, and assignment of US was based on participant's individual ratings; 2 = pretest, no individual assignment; 3 = no pretest, individual assignment; 4 = no pretest, no individual assignment	.90
CS duration	Duration of the CS presentation (in s)	Continuous	1.00
US duration	Duration of the US presentation (in s)	Continuous	1.00
CS supra-/subliminal	Was CS presented subliminally or supraliminally?	1 = supraliminal ($\geq$50 ms); 2 = subliminal (<50 ms)	—[a]
US supra-/subliminal	Was US presented subliminally or supraliminally?	1 = supraliminal ($\geq$50 ms); 2 = subliminal (<50 ms)	—[a]
A priori CS–US match	How were CS and US assigned to each other?	1 = yes, CS and US were matched according to one or another criterion (e.g., affective/perceptual similarity); 2 = no, there was random assignment/no a priori relation	.70
Same assignment	For paired trials, was a given CS always paired with the very same US?	1 = yes; 2 = no, a particular CS was paired with varying USs of the same valence	.89
(c) Nature of the response			
Type of dependent variable	Which dependent variable was used in order to assess the valence of the CS?	1 = self-reported liking; 2 = choice behavior; 3 = implicit measure; 4 = startle response magnitude	.92
Implicit measure	What was the specific type of implicit measure?	1 = affective priming; 2 = Implicit Association Test (IAT); 3 = Name Letter Task; 4 = Affective Misattribution Paradigm; 5 = Extrinsic Affective Simon Task	—[b]
CS Test	Was the evaluated CS at post-acquisition identical to the CS that had been paired during acquisition?	1 = yes, identical; 2 = not identical	.96
(d) Context			
Awareness/%aware (continuous)	Was there an assessment of contingency awareness (either inter-individually on the level of *ps* or intra-individually on the level of CS–US pairings)?	Contingency awareness assessment (1–5): 1 = yes, all *ps* in this (sub)group were classified as c-unaware; 2 = yes, all *ps* classified as c-aware; 3 = yes, all CS–US pairs in this analysis were classified as c-aware; 4 = all CS–US pairs were classified as c-unaware; 5 = yes, __% of *ps* in this (sub)group were classified as c-aware; 6 = no assessment	1.00

(*table continues*)

Table 2 (*continued*)

Variable	Definition	Coding options	IA
Learning	Was the experiment explicitly presented as a learning study?	1 = yes, 2 = no, it involved some kind of cover story	1.00
Spontaneity	Were participants urged to evaluate the CS in a spontaneous, intuitive, manner?	1 = yes, 2 = no	.94
	Abstract aspects (relation between stimuli)		
(a) Statistical properties of the relation			
Contingency (categorical)	What was the contingency between CS and US during acquisition?	1 = only partial; 2 = full, i.e., CS and US were paired on all trials	.96
No. CS only	Number of trials during acquisition in which a CS was shown without the US	Continuous	.96
No. US only	Number of trials during acquisition in which a US was shown without a CS	Continuous	1.00
Contingency index	$\frac{\text{No. paired trials}}{\text{No. paired trials} + \text{no. CS only} + \text{no. US only}}$	Continuous	—[a]
No. paired trials	Number of co-occurrences between CSs and USs during acquisition	Continuous	.94
(b) Temporal properties of the relation			
Presentation	What was the temporal sequence for CS–US pair presentations?	1 = forward (CS precedes US); 2 = backward (CS follows US); 3 = simultaneous: onset and offset were matched	.96
ISI	Time interval (in s) between CS and US presentation	Continuous	.80
ITI	Time interval (in s) between trials	Continuous	1.00
(c) (Dynamic) changes in the nature of the relation			
Special designs	Special design features that affect the nature of CS–US relation	I. Change in relation (1–6): 1 = latent inhibition (CS only trials before acquisition); 2 = extinction (CS only trials after acquisition); 3 = US pre-exposure; 4 = US postexposure; 5 = counterconditioning (pairing of already conditioned CS with a US of opposite valence); 6 = US revaluation (reversal of US valence after acquisition). II. Indirect relation (7–9): 7 = second-order conditioning (CS2 conditioned with US; then focal CS1 conditioned with CS2); 8 = sensory preconditioning (focal CS1 paired with CS2; then CS2 conditioned with US); 9 = occasion setting (Stimulus X is predictive of CS–US contingency)	—[b]

Note. IA = Interrater agreement (Cohen's kappa for categorical moderators and Pearson's *r* for continuous moderators).
[a] Codings for this category were computed from other codings in the database. [b] Codings for this category were made by Wilhelm Hofmann only.

presented, (c) the nature of the response that is observed, and (d) contextual features. For instance, studies may vary according to whether EC is studied in adults or children, the stimulus modalities of the CS and the US, whether liking is assessed directly or indirectly, and contextual features such as whether the study was explicitly presented as a learning study or involved a more covert presentation of the relation between stimuli.

For each moderator, we asked the general question of whether it explains significant amounts of variance in EC effect sizes across studies. By assessing the degree of generality of EC effects and identifying possible boundary conditions (or moderators), we hope to contribute to a solution of the three questions that have guided EC research: (a) Is EC a genuine and general phenomenon? (b) Is EC a unique form of Pavlovian conditioning? and (c) What are the mental processes that underlie EC? Information about the first question can be derived from our assessment of the overall magnitude and the impact of potential moderators. The identification of moderators will also provide information about the second and third questions.

Method

Literature Search Strategy

In our search, we focused on published or in-press articles, dissertations, book chapters, and unpublished manuscripts. The first search was conducted in February 2007 and updated continuously until December 2008. We retrieved published literature through a detailed search in PsycLIT and PsycINFO, the two main databases for psychological research articles, as well as Dissertation Abstracts, the main database for doctoral dissertations. We used the following six keywords: "evaluative conditioning," "evaluative learning," "affective conditioning," "affective learning," "affective conditioning," and "attitude learning." Our search was supplemented by hand searching the references cited in a number of major reviews and handbook chapters on evaluative conditioning as well as in a random sample of 30 published empirical articles. Because some studies in the domain of conditioned taste aversions may also qualify as (quasi-)experimental studies of evaluative conditioning (e.g., Rozin et al., 1998), we conducted an additional search using the key terms "taste aversion" and "food aversion." In order to locate gray literature (i.e., technical reports, unpublished manuscripts, articles currently in press), we sent e-mail requests to several cognitive, social, and personality psychology electronic mailing lists as well as to the participant list of a major 2007 international workshop on evaluative conditioning. Excluding nonhuman animal research and other clearly unrelated work such as case studies and biological research, this search strategy yielded a total of 286 unduplicated citations. Most of these (n = 282) could be retrieved either in electronic or print format for possible inclusion in our meta-analysis. As can be seen from the top box in Figure 1, the majority of citations consisted of published or in-press journal articles. However, there were also a substantial number of unpublished reports/drafts and doctoral dissertations.

Journal articles came from 79 different journals, with the highest number of articles published in *Learning and Motivation* (n = 26), *Behaviour Research and Therapy* (n = 19), *Cognition & Emotion* (n = 14), and *Journal of Personality and Social Psychology* (n = 13). According to the categories established by the ISI Web of

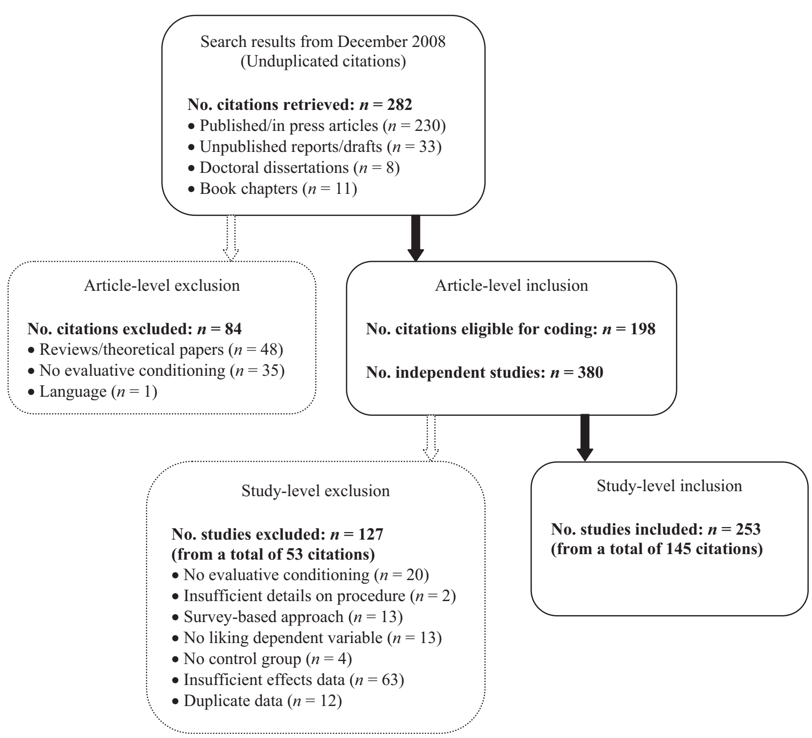


Figure 1. Quorum flowchart describing the sequence of steps and the criteria by which studies were included or excluded for the present meta-analysis.

Knowledge, journals were classified into the following broad categories: psychology (73%), neurosciences (8%), business (6%), psychiatry (5%), food science/nutrition (5%), communication (1%), education (1%), and multidisciplinary (1%). Of those journals in the psychology category, 27% were categorized as "multidisciplinary" (within psychology), 25% as "experimental," 15% as "clinical," 13% as "social," 10% as "biological," 5% as "applied," and 5% as "educational." The median impact factor of all journals was 2.4. The mean article publication year was 1993.

Across all citations in the database, the origin of the first author varied across 13 different countries from four continents with the following percentage distribution: United States of America (36.7%), Belgium (19.6%), United Kingdom (19.2%), Germany (8.5%), Australia (6.0%), the Netherlands (3.6%), Spain (1.8%), Canada (1.4%), Israel (1.0%), Italy (1.0%), Switzerland (0.4%), Japan (0.4%), and China (0.4%).

Study Eligibility

Of the 282 citations retrieved, 84 were excluded for one of the following reasons as determined by two independent judges (see Figure 1): On the basis of the title, the abstract, and a scanning of the text, the citation (a) was determined to be a review paper or otherwise theoretical in nature ($n = 48$), (b) clearly did not use EC procedures ($n = 35$), or (c) was written in a language not spoken by the authors (one Chinese citation). The remaining 198 citations yielded a total of 380 studies. The following five exclusion criteria were applied to determine the eligibility of each study for inclusion in the meta-analysis.

1. The study was an experimental or quasi-experimental instance of EC, whereby one (or several) CSs were paired with one (or several USs) of a given valence. Twenty studies were excluded because they did not include such pairings. These were typically pilot studies, thought experiments, or experiments focusing on other effects (e.g., mere exposure). Two further reports were excluded because they did not contain enough procedural information to allow for a definite judgment. Also, as part of the experimental or quasi-experimental pairing procedure, assignment of the CSs to the USs (and other parameters related to the pairing) should be under the control of the experimenter. We had to exclude 13 exploratory survey-based studies from the conditioned taste aversion domain in which CSs were not assigned to USs by the experimenter (and, consequently, most other pairing parameters varied widely and uncontrollably within studies). Rather, in those studies, respondents were asked to list any types of food they had consumed within a given time window (e.g., 24 hr) before a negative event and to indicate whether they had developed a taste aversion with regard to any of these food items.

2. In EC studies, it was examined whether the pairing of the CS with the US changed the valence of the CS. The dependent variable must therefore have provided an index of CS valence. In collecting data on changes in liking we did not want to be overly exclusive from the outset. We therefore included, along with traditional self-report measures (e.g., visual analog scales, semantic differentials), product choice, implicit measures of valence (e.g., affective priming, Implicit Association Test), and startle blink magnitude data. These are all determined to a large extent by stimulus valence (e.g., Vansteenwegen, Crombez, Baeyens, & Eelen, 1998). Skin conductance data were not included because skin conductance is generally considered an indicator of arousal rather than valence (Lang, 1995). Consumption data were not included because food intake is likely to be influenced by a host of other variables besides preferences, such as short- and long-term anticipated consequences, long-term goals, need states, and social/cultural factors (e.g., Herman & Polivy, 2004), which, taken together, render it a very distal measure of preference. Thirteen studies did not assess or report any of the eligible dependent outcome measures of liking or disliking after the acquisition stage and were therefore excluded.

3. The study design contained at least two out of the following three measurements (see also Figure 2 and text below): (a) the valence assessment of a CS after it was paired with a liked US (L); (b) the valence assessment of a CS after it was paired with a disliked US (D); and (c) a neutral comparison measure (N), such as the valence assessment of a CS that was paired with a neutral US, a CS that was not paired with a US, or a pre-acquisition valence assessment. Four studies had to be excluded because either only L or only D was realized within each study; thus, there was no second measure with which the valence of the CS could be compared in a meaningful way.

4. Studies reported the data in a way that at least one relevant effect size contrast could be coded from the data and transformed into *d* effect size statistics (see below). In many cases, for example, the relevant data were displayed in figures but the exact means and standard deviations were not provided. In addition, *F* tests from one-factorial tests involving more than two groups or more than one factor often were not analyzable by meta-analysis in the presented format. Great efforts were made

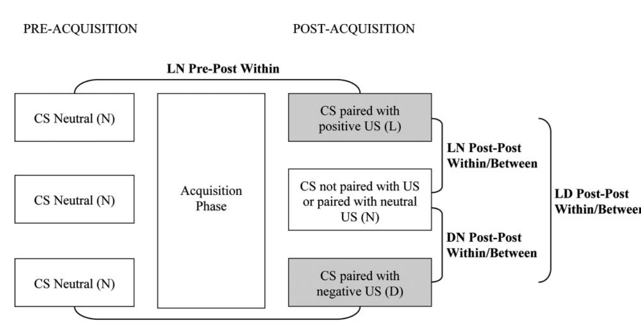


Figure 2. Illustration of possible effect size contrasts (boldface) in within- or between-subjects evaluative conditioning designs. CS = conditioned stimulus; US = unconditioned stimulus; L = "liking"; D = "disliking"; N = "neutral" (see text for details).

to contact authors by e-mail and request missing data for studies published less than 20 years ago. Even though a majority of requests (60%) were answered and the requested effect size data provided, effect size statistics could not be computed for a total of 63 studies.

5. Data were not included if they had already been reported in another citation included in the meta-analysis (n = 12). We established this criterion in order to avoid duplication.

After application of these exclusion criteria, 253 independent studies stemming from a total of 145 citations were retained for coding of study characteristics and effect sizes. The included citations are listed in the References section.[1]

Coding of Study Characteristics: Study-Level Data, Concrete Aspects, and Abstract Aspects

Eligible studies that were published in English or German were coded by two independent and trained psychology students, each of whom coded approximately half of the studies. The coding was done through the use of a data coding form and a clearly arranged coding manual. The coding form displayed all variables and possible coding options. The manual contained further brief explanations on the relevant coding variables and the respective category assignments. A random subset of 47 studies was coded separately by both coders in order to determine interrater agreement (see Table 2, right column). Before analyses, the codings for 75% of studies were double-checked by Wilhelm Hofmann. Cases with disagreement were resolved through discussion.

Each study was coded in a hierarchical manner that was also reflected in the arrangement of the coding sheet. Codings that could not be determined because the relevant information was either absent from the text or ambiguous were marked as missing. On the study level, we coded data referring to the whole study: study identification number (study-ID), authors, publication year, title, type of publication (e.g., journal article, dissertation), and journal. Studies also were classified according to the main research topic into one of the following six categories: general learning, consumer attitudes, social attitudes, fear, self-esteem, and other.

On the *procedural level*, data referring to the specific experimental implementation of the EC procedure were entered separately for each experimental (sub-)group reported in the study[2] and for each dependent variable assessing the degree of CS liking/disliking. Separate coding sheets were used for each experimental condition and for each dependent outcome measure of a given study. For instance, if a study included two experimental conditions and two different measures of liking, four coding sheets were used. The coded characteristics were organized according to the major distinction between concrete and abstract aspects of the EC procedure (see above). Table 2 provides the definitions and coding options for the concrete and abstract aspects from the coding manual. The concrete aspects were classified according to (a) the characteristics of the organism (e.g., sample composition), (b) the stimulus properties of the CSs and USs (e.g., CS/US modality), (c) the nature of the response that was observed (e.g., the type of dependent variable under investigation), and (d) contextual features (i.e., contingency awareness, learning context, instructions to judge CS spontaneously). Note that we classified the issue of contingency awareness as a contextual feature because it relates to the question of whether EC can be observed in a context where participants report awareness of the CS–US contingencies (see De Houwer, in press).

The abstract aspects concern the properties of the relation between CS and US that can be described without reference to concrete organisms, stimuli, responses, or contexts. As can be seen from Table 2, abstract aspects were divided into statistical properties of the relation (i.e., full vs. partial contingency, number of CS/US-only trials during acquisition, statistical contingency index), temporal properties of the CS–US relation (e.g., forward vs. backward vs. simultaneous presentation, interstimulus interval), and changes in the CS–US relation (i.e., the use of special designs such as extinction or latent inhibition). Table 2 also shows that interrater agreement computed from Cohen's kappa for categorical moderators and Pearson's r for continuous moderators was generally quite high (average interrater agreement = .93).

Coding of Effect Sizes

On the lowest level of the coding sheet, the available effect size data were coded according to a sophisticated effect data grid. The grid allowed for a convenient coding of measurement occasion for each mean (pretest, posttest, postextinction), effect size data (e.g., means, *SD*s, *N*s, t values), the direction of the effect size, and additional data such as the correlation between repeated measures for within-subjects data. All of these codings were highly reliable (average interrater agreement = .95). Again, the codings for 75% of studies were double-checked by Wilhelm Hofmann before analyses, and cases with disagreement were resolved through discussion.

Because the assessment of EC effects essentially involves the comparison of two means, we chose the standardized mean difference (Cohen's d) as the effect size statistic. One of the two means, the CS postacquisition score, is always given by the evaluation of a CS after it has been paired with a US of either a positive or a negative valence. We refer to these two possibilities as L (for the expected "liking" of the CS) and D (for "disliking"), respectively. These two means are illustrated by gray shading in Figure 2. Because EC study designs can take a number of forms, an EC effect can be computed in several ways depending on whether a within-subjects or a between-subjects design is chosen and on what type of CS evaluation is taken for comparison. These possibilities are illustrated in Figure 2. In a within-subjects design, evaluations provided by the same participants can be compared with each other. For instance, a comparison can be made between the mean of D (or L) and the respective evaluation of the supposedly neutral CS, denoted as N ("neutral"), *before* it has been paired with the US. The two resulting effect size contrasts can be called the *LN pre–post within* contrast and the *DN pre–post within* contrast, respectively (see Figure 2). Alternatively, in a within-

[1] A list of excluded citations may be obtained from Wilhelm Hofmann on request.

[2] Note that it was not necessary to code control groups from between-subjects designs on a separate sheet. The control means were entered at the effect-size level only.

subjects design, L or D can be compared with the postacquisition evaluation of a CS that has been presented but not paired with a US during the learning phase (*LN post–post within*; *DN post–post within*). Finally, L can be directly compared with D given that each participant was presented with a CS paired with a positive US and a CS paired with a negative US (*LD post—post within*).

In a between-subjects design, there is only one type of CS pairing per group of participants. Therefore, three effect size contrasts were possible: the *LD post–post between* comparison in designs where the CS was paired with a positive US in one group and a negative US in another group, the *LN post–post between* comparison, and the *DN post–post between* comparison, the latter two of which involved evaluations of a CS paired with a positive or negative US, respectively, in one group with evaluations of the same CS in a neutral control group (see Figure 2). The nature of the neutral control group was coded according to whether the CS of interest (a) has been paired with a neutral US (11%), (b) has been presented during acquisition without a US (22%), (c) has been randomly paired with USs of varying valence (48%), (d) had been rated prior to the acquisition phase in the control group (8%), or (e) other cases, such as studies where two different control groups were merged for comparison purposes (11%).

Because we were interested in whether EC effects vary as a function of the specific effect size contrast used, the exact type of contrast was specified for all effect size codings on the coding grid as was the nature of the neutral control group in between-subjects designs. All possible contrasts that could be computed on the basis of the reported or provided data for each study were entered into our database with the following exception: Data from subgroups not in our focus of interest (e.g., Black vs. White participants) were only entered if no overall report of effect size (i.e., collapsed across subgroups) was provided.

Effect-size transformations. Effect sizes were preferably computed from group means, standard deviations (or standard errors converted to standard deviations), and *N* per group. Cohen's *d* could be computed from this kind of data in 66.9% of cases. Additionally, we transformed effect sizes into *d*s from *t*-test statistics (8.9%), *F* statistics involving one–degree of freedom tests and only two groups (15.1%), mean gain scores (3.8%), proportion differences (χ^2; 2.3%), or other effect size data, such as Pearson's *r* (3.0%), using the appropriate formulas (e.g., Lipsey & Wilson, 2001; D. Rosenthal, 1991).

Because *d* computed from means, standard deviations, and sample sizes is the appropriate standardized level of effect regardless of whether the data stem from a between- or a within-subjects design (Dunlap, Cortina, Valsow, & Burke, 1996), effect sizes from both designs are directly comparable to each other. However, in the case of within-subjects designs, transformation from *t* statistics, *F* statistics, and mean gain scores tend to overestimate *d* to the extent that there is a correlation between repeated measures (Dunlap et al., 1996). In order to correct for this bias in these latter cases, we either (a) recorded the correlation between measures in all cases where it was provided or (b) calculated the correlation from cases in our data set where *d* could be computed both from means, standard deviations, and sample sizes (resulting in unbiased *d*s) as well as from *t*, *F*, or mean gain data (resulting in biased *d*s). The average within-subjects correlation based on these 65 data points was moderate ($r = .35$). In cases where the within-subjects correlation was not provided or could not be computed, we imputed the above-average within-subjects correlation. All effect sizes involving biased *d* data were then corrected downwards with the appropriate formulas in order to render all within- and between-design effect sizes comparable for the whole data set (e.g., Dunlap et al., 1996).

Dependent outcome measures. The coding of effect sizes for self-report measures and most other dependent outcome measures was straightforward. With regard to affective priming measures, we coded only reaction time data (but not error rate data) according to the following rationale: Data were preferably entered as the effect size contrast between affectively incongruent and affectively congruent combinations of CS primes and target stimuli (with relatively faster reaction times to congruent combinations indicating a positive EC effect). The majority of affective priming studies ($n = 23$; 74%) reported effect size data this way. For six additional studies, the affective priming effect of interest was reported equivalently as the interaction effect of a two-way analysis of variance (ANOVA) with CS (positive vs. negative) and target (positive vs. negative) as variables. In this case, we converted the *F* value of the interaction to *d* using the appropriate formula (e.g., Sedlmeier & Renkewitz, 2007). For three remaining studies, reaction times for affectively incongruent and affectively congruent trials were available only separately for positive and negative CSs. For two of these studies, both the CS prime and the target main effects were virtually zero (and all *F*s < 1), so these simple main effects could be averaged to properly estimate the overall congruent–incongruent contrast. For the third study, there was a main effect of CS prime, so the data could not be included, as averaging the simple main effects would have resulted in biased effect size estimates (i.e., confounded by the main effect). With regard to startle response data, priority was given to startle response magnitude as an indicator of valence (where stronger startle response magnitude is indicative of a more negative valence with regard to the CS). Startle response latency was not coded.

Meta-Analytic Procedure

Effect-size correction formulas, standard errors, and weights. As recommended by Lipsey and Wilson (2001), we adjusted effect sizes for bias in small samples using the correction formula proposed by Hedges (1981). We then computed standard errors for unbiased effect sizes using the appropriate formulas detailed in Lipsey and Wilson (2001). The inverse of the squared standard error was used as weights for the meta-analysis whereby high-precision effect-size estimates gain more weight than do low-precision estimates.

Combination of multiple effect sizes within studies. Because many studies used more than one experimental procedural manipulation, included more than one dependent variable, and/or allowed for multiple effect size codings per dependent variable, most primary studies yielded more than one effect size. In order to ensure independence of the effect sizes entered into the meta-analysis (Lipsey & Wilson, 2001), we performed a two-stage procedure. In the first stage, we selected for each type of (moderator) analysis the relevant effect sizes from the data set and aggregated (i.e., averaged) multiple effect sizes and their weights within studies (Hedges & Olkin, 1985). All subsequent meta-analytic computations were then performed on the aggregated study effect sizes. This two-stage procedure was repeated for all runs of the meta-analysis such that different sets of *d*s within studies were averaged, depending on the analysis in question. For instance, for the overall analysis of self-report measures, all *d*s pertaining to self-reported data were averaged within each study first before being submitted to the overall analysis. For the moderator analyses, we aggregated all *d*s pertaining to a particular moderator category within each study and then submitted the aggregated data to an ANOVA or regression analysis. Generally, moderator categories varied across but not within studies, so a study typically provided information for one value of the moderator. In cases where a moderator varied also within a given study (e.g., visual and odor US modalities used within one study), we allowed the study to provide aggregated information for each value of the moderator for which information was given (i.e., visual and odor) in order to make full use of the data at hand.

Treatment of special designs and reduction of redundancies. For the preliminary analysis, overall analysis, and moderator analyses, except where noted, we excluded effects from special designs, such as latent inhibition or EC effects assessed after an extinction phase. Additionally, in order to reduce redundancies among data entries, we excluded redundant subgroups and entered only the effect size that collapsed the data across these subgroups unless the subgroups were in the focus of interest (e.g., for the analysis of contingency-aware vs. contingency-unaware subgroups). Moreover, if both pre–post and post–post analyses were provided in a within-design, we entered only the post–post comparisons because the post–post comparison can be considered to be the superior control group in that it also controls for mere exposure. In case multiple comparison groups were provided in a between-design, we entered only one comparison group in the following order of priority: (a) CS paired with a neutral US, (b) CS presented without a US, (c) CS randomly paired with USs, (d) pre-acquisition rating, and (e) other.

Meta-analytic computations. For the overall analysis and all subsequent analyses, we chose a mixed-effects model, that is, a special type of random-effects model (Lipsey & Wilson, 2001). The mixed model assumes that there is variation in effect sizes beyond sampling error that can be attributed partly to systematic factors (i.e., coded study characteristics) and partly to unmeasured (and possibly unmeasurable) random sources (Lipsey & Wilson, 2001). For three reasons, this model was advocated over the fixed-effects model, which posits that all variations in effect sizes are attributable only to sampling error and not to true variation on the level of population effect sizes. First, given the large number of parameters involved in the EC procedure and the degree of debate about EC effects in the literature, the assumption of a fixed-effects model is quite unrealistic, and thus a random-effects model should be advocated (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009). Second, because the mixed model assumes a random component in addition to sampling error, the confidence intervals estimated under this model will be larger than under a fixed-effects model. This renders the mixed-effects model the more conservative approach to detect true variation in effect sizes (both for the overall heterogeneity analysis and for moderator analysis). In contrast to the mixed-effects model, the fixed-effects model has been criticized for high Type I error rates (Overton, 1998; Schmidt, Oh, & Hayes, 2009). Third, the mixed-effects model converges on the fixed-effects model in the unlikely case that the random between-variance component is zero. The mixed-effects model can therefore fully replace the fixed-effects model in case the assumptions of the latter are fulfilled (but not vice versa).

The mixed-effects model was implemented through a set of three SPSS macros developed by David Wilson (see Lipsey & Wilson, 2001) for the overall analysis, the categorical moderator analyses, and the weighted regression analyses. In all analyses, study effect sizes were weighted by their inverse variance (e.g., Lipsey & Wilson, 2001). The random-effects variance component was based on the method of moments estimation. We applied Cochran's Q test of heterogeneity and the percentage-based I^2 measure of heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003) to judge the degree of heterogeneity in effect sizes.

To address whether variations in effect sizes can (at least partly) be explained by the coded study characteristics in question, we performed meta-analytic ANOVAs for categorical moderator variables and weighted least squares regression analyses for continuous moderators (Lipsey & Wilson, 2001). In the meta-analytic ANOVA analog, variance in study effect sizes is partitioned into the portion explained by the categorical variable (Q_B) as an indicator of variability between group means and the residual remaining portion (Q_W) as an indicator of variability within groups. Q_B is tested for significance against a chi-square distribution with $df = j - 1$ (where j is the number of categories or groups). A significant between-groups effect indicates that the variance in effect sizes is at least partially explained by the moderator variable. Moderator groups for which fewer than three cases were available were not included in these analyses. If the moderator analysis involved more than two valid groups, simple contrasts following the procedure by R. Rosenthal and Rubin (1982) were applied in order to determine significant differences between groups.[3]

To assess the relationship between continuous moderators and study effect sizes, we used the weighted regression analysis macro. The macro contains a built-in correction of the standard errors for the proper estimation of the standardized regression coefficients and significance levels (Hedges, 1994; Lipsey & Wilson, 2001).

Results

Overview

Before conducting our main sets of analyses, we conducted two types of preliminary analyses on the full set of 253 included studies in order to safeguard against two potential dangers in

[3] Linear contrast weights for the compared categories were +1 and −1, and the conservative inverse of the variance estimates from the random-effects model were used for computing the contrast between category means (see also Rosenthal, 1991). Note also that the test of significance of group differences through contrasts (or *F* test in the case of only two categories) is not identical to the test of overlap of confidence intervals around the respective means. That is, even in the case of confidence interval overlap, between-group differences can nevertheless be statistically significant (Cumming & Finch, 2005; Estes, 1997). This is because the between-groups test is based on a joint estimate of the standard error of the difference between means, whereas confidence intervals around a single mean do not reflect between-groups information. The confidence intervals therefore should be best interpreted as providing valuable information about the precision of single estimates and their difference from zero (Rouder & Morey, 2005).

meta-analysis. First, we wanted to assess whether the present meta-analytic sample might have been affected by publication bias. A second danger is that actually different things are compared that should be kept separate from another (the "apples and oranges" problem). In order to judge whether effect sizes from different designs and use of different outcome measures are comparable to each other or should instead be treated separately, we conducted a set of preliminary moderator analyses in order to determine the main sample of analysis. After presentation of the results for this main sample, we report the results of a sensitivity analysis assessing the degree to which our conclusions are affected by different selection criteria for effect size inclusion in the meta-analysis.

Preliminary Analyses

Assessment of potential publication bias. A possible danger to the validity of any meta-analysis is the presence of publication bias against nonsignificant findings (e.g., Borenstein et al., 2009). This so-called "file-drawer problem" (R. Rosenthal, 1979) usually leads to an overestimation of effect sizes. We first used a funnel plot, that is, a plot of sample size versus effect size, in order to assess whether potential bias was present (Light & Pillemer, 1984). A publication bias against nonsignificant findings implies that only large effects are reported by studies with small sample sizes, as only large effects reach statistical significance in small samples. Thus, a publication bias should manifest itself graphically in a cutoff of small effects for studies with small sample size. From the funnel plot, an exclusion of null results was not apparent because many small or even negative effect sizes were reported by small-sample studies (see Figure 3). Second, we statistically evaluated the relationship between effect size and sample size with the Egger test (Egger, Smith, Schneider, & Minder, 1997), a regression of reported effect sizes on sample size. The standardized regression weight was close to and not significantly different from zero ($\beta = -.029$, $p = .377$), indicating that sample size and effect size were not confounded. Drawing on these findings, we conclude that a publication bias seems unlikely for the present meta-analysis.

Selection of main analysis sample. We conducted a number of preliminary ANOVAs on the full set of 253 included studies using the method described above in order to determine whether some of the general design-related aspects, such as type of effect size contrast (LD, LN, DN), type of design (within vs. between), measurement occasion (pre–post vs. post–post), type of control group, research topic, and type of dependent outcome measure (self-report, choice, implicit measures, eyeblink startle response data) had a general influence on the magnitude of effect sizes. First, effect sizes from within-subjects designs ($d = .48$, $SE = .027$) were somewhat larger than those from between-subjects designs ($d = .39$, $SE = .043$), but the difference was not statistically significant, $Q_B(1) = 3.34$, $p = .067$, suggesting that both types of effect sizes can be analyzed together. Second, as expected, effect sizes from LD contrasts ($d = .47$, $SE = .034$) were somewhat larger than those from the LN ($d = .40$, $SE = .035$) and DN ($d = .46$, $SE = .040$) neutral comparisons, but the difference was not statistically significant, $Q_B(2) = 2.23$, $p = .328$. We therefore decided to include LD contrasts alongside the other two contrasts in the main analyses, also because dropping this contrast would have meant discarding information from a substantial number of studies ($n = 83$) that employed or reported only LD contrasts. Third, effect sizes from pre–post contrasts ($d = .46$, $SE = .043$) were comparable in magnitude to effect sizes from post–post contrasts ($d = .46$, $SE = .026$), $Q_B(1) = 0.01$, $p = .925$. Fourth, effect sizes from between-subjects designs did not vary systematically as a function of which type of neutral control group was used, $Q_B(4) = 0.79$, $p = .939$. Fifth, effect sizes were comparable across the different research topics under investigation, $Q_B(5) = 6.35$, $p = .274$.[4] However, type of dependent outcome measure yielded a strong moderator effect, $Q_B(3) = 17.09$, $p = .001$, indicating that EC effect sizes varied as a function of the specific outcome measure used. Because the vast majority of dependent variables (79%) in the database were self-report measures, we decided to focus on this type of measure as the primary outcome of interest for the main analysis. A more fine-grained analysis involving the remaining dependent outcome measures is provided in an additional section later in the article.

These preliminary analyses are reassuring in that more general design-, effect-size-, and topic-related characteristics did not exert a general influence on the magnitude of effect sizes in our data set, with the exception of type of dependent outcome measure. Therefore, the main overall analysis as well as the initial moderator analyses of study characteristics was conducted on the set of self-report measures, including both within- and between-subjects data and all types of effect sizes contrasts. In order to safeguard against possible outliers (e.g., Lipsey & Wilson, 2001; Nelson & Kennedy, 2009), we applied a threshold of ± 3 *SD* for the overall analyses as well as for all subsequent moderator and regression analyses.

Overall Analysis: Average Effect and Heterogeneity Assessment

We now turn to the estimation of the overall effect size across studies in the main sample of analysis and to the assessment of heterogeneity in effect sizes. The overall analysis to address these issues was conducted on a total of 215 study effect sizes on self-report outcome measures, stemming from 652 coded primary effect sizes within studies. One study effect size ($d = 2.483$) was excluded as an outlier (as it was greater than ± 3 *SD*s away from the sample mean). The remaining 214 study effect sizes were computed from a total of 9,149 participants. The random-effects model yielded a mean estimated effect size (d) of .524 and a standard error (*SE*) of .030. According to convention (J. Cohen, 1977), this can be considered a medium average effect. The 95% confidence interval around this estimate had a lower limit of .466 and an upper limit of .582. The effect was significantly different from zero ($Z = 17.76$, $p < .001$). The minimum and maximum study effect sizes were -0.805 and 1.920, respectively. Cochran's Q statistic yielded a significant effect, $Q(213) = 706.97$, $p < .001$, indicating heterogeneity. The estimated random variance component, that is, the portion of the total variance ($V_{tot} = 0.170$) attributable to true variation on the level of population effects

[4] Specifically, $d_{learning} = .52$ ($SE = .04$), $d_{consumer\ attitudes} = .47$ ($SE = .04$), $d_{social\ attitudes} = .46$ ($SE = .06$), $d_{fear} = .49$ ($SE = .10$), $d_{self\text{-}esteem} = .22$ ($SE = .12$), $d_{other} = .47$ ($SE = .14$).

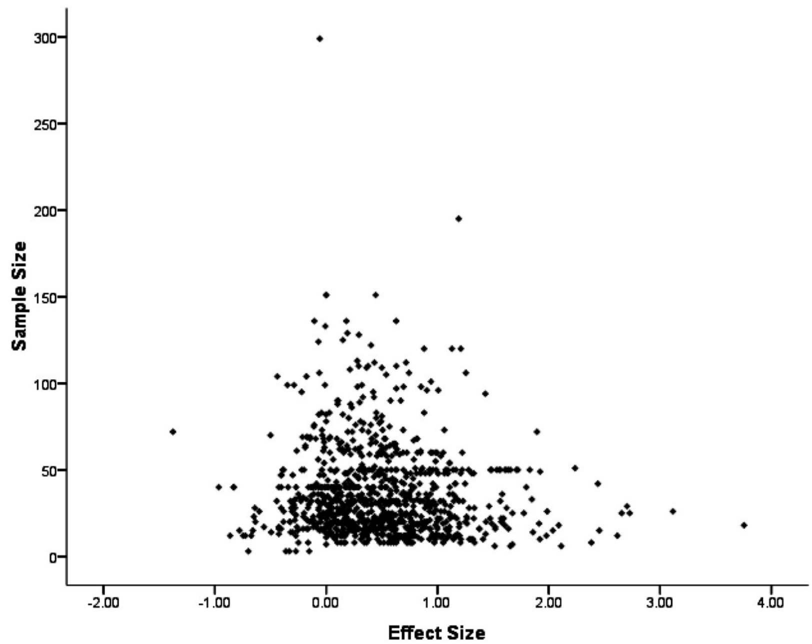


Figure 3. Scatterplot of effect size against sample size ("funnel plot"; Light & Pillemer, 1984) as a visual aid to detect the presence of publication bias. If the direction of the effect is toward the right (as in our meta-analysis), publication bias against null results should manifest itself as a lack of small-sample studies reporting small effects.

was considerable (V = .120). Correspondingly, the I^2 statistic (Higgins et al., 2003) indicated that 70% of the variance in effect sizes across studies was indicative of true heterogeneity of effect sizes, justifying the choice of a random-effects model.

Moderator Analyses

The next major issue addressed was the identification of moderators that can account for significant portions of the (large) variation in effect sizes across studies. In presenting these results, we follow the aforementioned organization scheme dividing moderator candidates into concrete and abstract aspects of the EC procedure. The main results from these analyses are summarized in Tables 3 and 4, respectively.

Concrete Aspects of the Procedure

Organism. The meta-analytic ANOVA yielded a significant effect for the sample under investigation. As can be seen from Table 3, effect sizes were comparable in magnitude for psychology students, other students, mixed samples (students and nonstudents), nonstudent samples, and samples suffering from psychopathology, with the latter showing the largest EC effects on a descriptive level. However, as indicated by the contrast indices in Table 3, EC effects in the nine studies involving children as participants were markedly smaller than in studies with psychology students, other students, nonstudents, and pathological samples. Moreover, the mean effect for children did not differ significantly from zero (as indicated by the inclusion of zero in the 95% confidence interval). Sample accounted for a total of 6.3% of the variation in effect sizes. With regard to the gender composition of the sample, a weighted least squares regression analysis on effect sizes as a function of the proportion of women in each sample yielded a nonsignificant standardized regression weight close to zero (see Table 3). Hence, EC effect sizes were largely unaffected by gender.

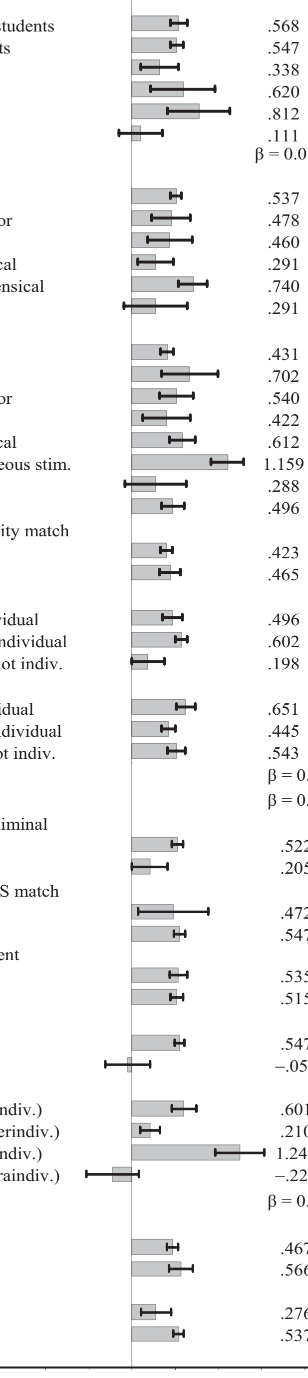
Stimulus properties. We now turn to properties of the CS and US, such as their modality, selection, and duration. With regard to CS modality, a significant moderator effect emerged (see Table 3). An inspection of contrast among effect sizes showed that EC effects were of a similar magnitude for visual, taste/flavor, and odor CS. Effect sizes were smaller for sensical verbal material and haptic stimuli. Both of these mean effect sizes were significantly different from the mean effect for nonsensical verbal material, which yielded the largest EC effects.

US modality had a significant impact on the magnitude of EC effects. As can be seen from Table 3, EC effects were most pronounced for studies involving electrocutaneous stimulation (i.e., mild electric shock) as the US. EC effects obtained with electrocutaneous stimulation were significantly different from EC effects with USs of other modalities. In contrast, haptic material (i.e., different textures that had to be touched) was associated with relatively small EC effects. Comparable effects emerged for visual, auditory, taste/flavor, sensical verbal stimuli, and "other" US modalities (which were mostly combinations of the former modalities, such as words and images presented together).

In addition to treating CS and US modality separately, we investigated whether EC effects differed as a function of whether CS and US were of the same modality or of different modalities. To this end, we created a new variable, CS–US modality match, by assigning a value of 1 to cases where CS and US modalities were identical and a value of 2 where modalities differed. Additionally, in order to guarantee a symmetric nature of this index, we included

Table 3
Moderator Analyses for Categorical and Continuous Moderator Variables: Concrete Aspects

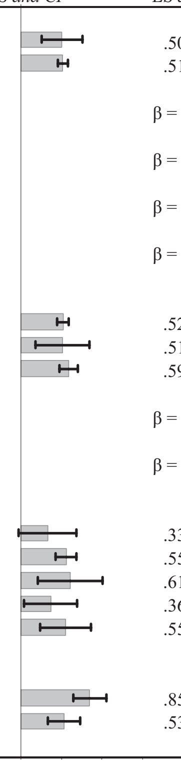
Moderator	ES and CI	ES *d*/β	*C*	*SE*	*K*	Q_B (*df*)	Q_W (*df*)	R^2	*p*	Level
Sample						16.01 (5)	238.44 (208)	6.3%	.007	**
Psychology students		.568	a	.051	69					
Other students		.547	a	.041	112					
Mixed		.338	a,b	.116	13					
Nonstudents		.620	a	.199	5					
Pathological		.812	a	.191	6					
Children		.111	b,c	.132	9					
Gender		β = 0.011		.170	117	0.02 (1)	135.4 (115)	0.0%	.895	
CS modality						12.55 (5)	240.05 (209)	5.0%	.028	*
Visual		.537	a,d	.035	146					
Taste or flavor		.478	a,b,c,d	.117	13					
Odor		.460	a,b,c,d	.138	10					
Verbal sensical		.291	b,d	.110	15					
Verbal nonsensical		.740	c	.087	24					
Haptic		.291	a,b,d	.194	7					
US modality						48.31 (7)	228.22 (211)	17.5%	.000	**
Visual		.431	a,b,e	.040	106					
Auditory		.702	a,b,c	.176	6					
Taste or flavor		.540	a,b,c,e	.103	16					
Odor		.422	a,b,c,e	.148	8					
Verbal sensical		.612	b,c,e	.081	27					
Electrocutaneous stim.		1.159	d	.102	17					
Haptic		.288	a,c,e	.188	7					
Other		.496	a,b,c,e	.069	32					
CS–US modality match						0.33 (1)	150.3 (146)	0.2%	.565	
Unimodal		.423	a	.036	113					
Cross-modal		.465	a	.064	35					
CS selection						15.11 (2)	215.49 (188)	6.6%	.001	**
Neutral/individual		.496	a	.062	49					
Neutral/not individual		.602	a	.036	125					
Not neutral/not indiv.		.198	b	.100	17					
US selection						8.55 (2)	237.56 (211)	3.5%	.014	**
Pretest/individual		.651	a	.058	61					
Pretest/not individual		.445	b	.042	100					
No pretest/not indiv.		.543	a,b	.057	53					
CS duration		β = 0.071		.000	192	1.11 (1)	217.32 (190)	0.5%	.292	
US duration		β = 0.082		.000	185	1.44 (1)	213.32 (183)	0.7%	.230	
US supra-/subliminal						8.01 (1)	203.79 (176)	6.6%	.005	**
Supraliminal		.522	a	.034	163					
Subliminal		.205	b	.107	15					
A priori CS–US match						0.13 (1)	207.14 (179)	0.1%	.721	
Yes		.472	a	.207	5					
No		.547	a	.033	176					
Same assignment						0.09 (1)	237.94 (206)	0.0%	.763	
No		.535	a	.052	69					
Yes		.515	a	.037	139					
CS test						19.43	243.86 (213)	7.4%	.000	**
Identical		.547	a	.029	207					
Not identical		−.053	b	.133	8					
Awareness						68.98	120.47 (96)	36.4%	.000	**
Aware (interindiv.)		.601	a	.073	37					
Unaware (interindiv.)		.210	b	.058	49					
Aware (intraindiv.)		1.245	c	.144	7					
Unaware (intraindiv.)		−.226	d	.155	7					
Aware %		β = 0.373		.003	85	13.7 (1)	85.03 (83)	13.9%	.000	**
Learning						1.67 (1)	224.96 (198)	0.7%	.197	
Cover		.467	a	.033	161					
Explicit		.566	a	.069	39					
Spontaneity						7.79 (1)	232.86 (207)	3.2%	.005	**
Yes		.276	a	.088	21					
No		.537	b	.030	188					



Note. CS = conditioned stimulus; US = unconditioned stimulus; stim. = stimulation; indiv. = individual; ES *d* = effect size estimate (Cohen's *d*); β = standardized regression coefficient; C = contrast index: different subscripts indicate significant differences ($p < .05$), as indicated by contrasts (Rosenthal & Rubin, 1982); *SE* = standard error; *K* = number of study effect sizes for a given moderator category/continuous predictor; Q_B = analysis of variance (ANOVA) between groups/regression sum of squares (*df*s); Q_w = ANOVA/regression sum-of-squares residual (*df*s); R^2 = squared multiple correlation indicating the proportion of variance between studies explained by a given moderator; *p* = significance level of between-groups effect (ANOVA) or regression coefficient.
[*] $p < .05$. [**] $p < .01$.

Table 4
Moderator Analyses for Categorical and Continuous Moderator Variables: Abstract Aspects

Moderator	*ES and CI*	*ES d*/ β	*C*	*SE*	*K*	Q_B (*df*)	Q_W (*df*)	R^2	*p*	Level
Contingency (categorical)						0.00 (1)	241.73 (212)	0.0%	.939	
Only partial		.508	a	.130	12					
Full		.518	a	.030	202					
No. CS only		β = 0.007		.016	219	0.01 (1)	247.71 (217)	0.0%	.915	
No. US only		β = −0.074		.005	220	1.38 (1)	248.32 (218)	0.6%	.240	
Contingency index		β = 0.045		.206	205	0.48 (1)	232.65 (203)	0.2%	.489	
No. paired trials		β = 0.054		.003	213	0.71 (1)	239.40 (164)	0.3%	.401	
Presentation						1.10 (2)	228.64 (200)	0.5%	.576	
Forward		.520	a	.036	146					
Backward		.515	a	.169	7					
Simultaneous		.592	a	.059	50					
Interstimulus interval		β = −0.009		.002	179	0.02 (1)	200.65 (177)	0.0%	.884	
Intertrial interval		β = 0.004		.001	157	0.00 (1)	175.31 (155)	0.0%	.957	
Special designs						2.44 (4)	28.26 (29)	7.9%	.656	
Latent inhibition		.333	a	.182	3					
Extinction		.558	a	.065	20					
US pre-exposure		.611	a	.203	3					
Second-order cond.		.369	a	.168	3					
Sensory preconditioning		.552	a	.161	5					
Extinction						4.82 (1)	39.46 (36)	10.9%	.028	*
Postacquisition effect		.851	a	.104	19					
Postextinction effect		.533	b	.101	19					



Note. cond. = conditioning; CS = conditioned stimulus; US = unconditioned stimulus; ES *d* = effect size estimate (Cohen's *d*); β = standardized regression coefficient; C = contrast index: different subscripts indicate significant differences ($p < .05$) as indicated by contrasts (Rosenthal & Rubin, 1982); *SE* = standard error; *K* = number of study effect sizes for a given moderator category/continuous predictor; Q_B = analysis of variance (ANOVA) between groups/regression sum of squares (*df*s); Q_w = ANOVA/regression sum-of-squares residual (*df*s); R^2 = squared multiple correlation indicating the proportion of variance between studies explained by a given moderator; *p* = significance level of between-groups effect (ANOVA) or regression coefficient.
$^{*} p < .05$. $^{**} p < .01$.

only those modalities in the computations that were present for both US and CS in the data set (visual, taste/flavor, odor, verbal sensical, haptic) and excluded those modalities that were exclusively used as either CS or US (auditory, verbal nonsensical, electrodermal) as well as modalities assigned to the category "other." As can be seen from Table 3, unimodal and cross-modal EC effects were not significantly different from each other.

Next, we investigated the ways in which stimuli were chosen (CS selection/US selection). As can be seen from Table 3, it did not matter whether CSs were individually assigned to participants (on the basis of pretest scores) or assigned on a group level as long as it had been ascertained that CSs were evaluatively neutral. However, both effect sizes differed markedly from EC effects stemming from studies in which the CS had initial valence and was assigned on a group level.[5] Thus, EC was clearly more successful when evaluatively neutral as compared with evaluatively significant CSs were used.

In a similar vein, US selection exerted a significant effect. EC effects were larger when USs were assigned individually to participants on the basis of participants' pretest scores than when no individual assignment based on pretesting was used (see Table 3). Somewhat unexpectedly, the effect for the third category—unpretested and not individually assigned USs—did not differ from the former two effects.[6]

Next, we scrutinized CS and US presentation time in more detail. First, we treated CS duration and US duration as continuous predictors. As can be seen from Table 3, both regression coefficients were positive in magnitude but did not differ significantly from zero. In a follow-up analysis, we contrasted supraliminal CS or US presentations with subliminal presentations, defined as presentation times of less than 50 ms. With regard to CS presen-

[5] The fourth possible category—not neutral, individually assigned—was empty because, quite naturally, there were no studies that assigned valenced CSs to participants on the basis of pretest values.

[6] As expected on logical grounds, the fourth possible category—no pretest, individually assigned—was empty.

tation, there were not enough studies using subliminal CS presentations in the context of self-report outcome measures to warrant a moderator analysis (but see the additional analysis on all dependent variables below). With regard to US presentation, a substantial number of studies ($n = 15$) used subliminal US presentations, with a median US duration of 17 ms (minimum = 2 ms; maximum = 29 ms). The moderator analysis showed that EC effects involving subliminal US presentation were significantly smaller than EC effects involving supraliminal US presentation (see Table 3). Moreover, the EC effect for subliminal US presentations was not reliably greater than zero as indicated by its confidence interval.

Finally, no reliable differences emerged when we compared studies regarding whether there was an a priori match between CS and US (a priori CS–US match). Specifically, effects were comparable in magnitude when a CS and a US of a given valence were matched according to some criterion, such as perceptual similarity, as compared with a random assignment of CS and US into pairs (see Table 3). Furthermore, it did not matter whether the same specific CS and US were always assigned together (same assignment) or whether a given CS was paired with varying specific USs of the same valence during acquisition.

Response properties. Next, we investigated properties of the response. Note that in this analysis we included only self-report measures and thus did not address the question of whether EC effects differ as a function of the type of measure of valence. Effects concerning the type of outcome measure are treated in more detail in the section entitled "Additional Findings Involving All Dependent Outcome Measures." Here, we address only one question related to the stimulus specificity of EC effects (CS test): Do EC effects for the trained CS transfer to other stimuli that are similar to the CS but have not been presented during the learning phase? For instance, the unpaired test stimulus may be a different exemplar from the category to which the CS belongs, or it may share some properties (e.g., shape) while being different on others (e.g., color). As reported in Table 3, EC effects were considerably reduced and not significantly different from zero when an unpaired test stimulus was used that resembled the CS. Note, however, that the number of study effect sizes for nonidentical CSs was relatively small.

Context. As a final concrete procedural aspect, we took into account three contextual features: contingency awareness, the explicitness of the learning context, and whether participants were urged to make their evaluation in a spontaneous, intuitive manner. First, we investigated whether EC is dependent on contingency awareness, that is, whether EC occurs only in a context in which the participant is aware of the CS–US contingencies. We addressed this issue by comparing EC effects for aware as compared with unaware samples or subsamples of participants (i.e., interindividual approach) as well as for aware as compared with unaware CS–US pairings (i.e., intraindividual approach) in a single analysis. As can be seen from Table 3, EC effects were almost three times as large for participants classified as contingency aware than for contingency-unaware participants, and this difference was significant. Nevertheless, the mean effect across the 49 samples of unaware participants was still reliably greater than zero. The aware–unaware difference became even more pronounced when considering the seven studies that used an intra-individual approach to studying the effect of contingency awareness (e.g., Pleyers et al., 2007). Here, the average EC effect was very large when taking into account only CSs from CS–US pairs of which participants were aware, but EC was absent and even slightly negative for CSs from CS–US pairs of which participants were unaware. In fact, the latter two values were the maximum and minimum estimated means in the whole moderator analysis. Overall, contingency awareness had a significant impact as a moderator and explained about 36% of the variance in effect sizes in that analysis.[7] Finally, in order to include those studies where an awareness assessment was implemented but no subgroup data were available (e.g., "61% of all participants were aware of contingencies"), we treated the degree of contingency awareness as a continuous predictor of EC effects in a regression analysis. The regression analysis confirmed that EC effect sizes increased with increasing degrees of contingency awareness (see Table 3).

Second, with regard to the learning context, we found that EC effects were descriptively larger for studies that made the learning situation explicit than for studies using some kind of cover story, but the difference was not statistically significant. Third, with regard to spontaneity, we observed that EC effects were significantly reduced when participants were urged to evaluate the CS in a spontaneous manner as compared with a default CS evaluation context in which no such instructions were given.

Abstract Aspects of the Procedure

We now turn to those moderators that relate to the abstract core of the regularity between CS and US. These aspects concern the nature of the relation between the CS and the US independently of the specific stimuli, responses, organisms, and contexts that are used to implement the CS–US relation. Results for these moderators are provided in Table 4.

Statistical properties of the CS–US relation. Effect sizes varied only trivially as a function of whether the statistical contingency between the CS and the US was partial (i.e., less than one) or complete (i.e., equal to one). In a similar vein, the number of CS-only and US-only trials during the acquisition phase did not contribute significantly to the prediction of EC effect sizes. Furthermore, a more sophisticated continuous index of statistical contingency—computed as the number of paired trials divided by the sum of paired trials, CS-only, and US-only trials—did not yield a significant moderator effect even though a slight positive trend emerged (see Table 4). Thus, EC effects appeared to be relatively robust with regard to deviations from perfect contingency between CS and US presentations. Furthermore, we investigated whether the absolute number of trials on which CS and US are paired (i.e., "co-occurrence") moderates EC effect sizes across studies. Even though the regression coefficient was in the expected

[7] A comparable ANOVA effect of contingency awareness was obtained when we additionally categorized participants from studies involving subliminal US or CS presentations as unaware of CS–US contingencies because subliminally presented stimuli are typically not noted and, hence, participants presumably remain unaware of contingencies, $Q_B(3) = 78.21$, $p < .001$, $R^2 = 37.0\%$.

positive direction, it did not reach statistical significance (see Table 4).[8]

Temporal properties of the CS–US relation. Do EC effects hinge on whether, during acquisition, the CS precedes the US (forward conditioning), the CS follows the US (backward conditioning), or onset of the CS and the US occur simultaneously? On a descriptive level, EC effects were slightly larger for the simultaneous presentation than for forward or backward conditioning. These differences were far from significant, however, indicating that EC effects appear to be relatively robust to this procedural parameter. Moreover, the absolute interstimulus interval between CS and US onset for each paired trial as well as the absolute intertrial interval between CS–US pairings did not moderate effect sizes, as indicated by the very low standardized regression weights for these continuous moderators (see Table 4).

Special EC designs. Finally, we extended our meta-analysis to special designs that go beyond the basic EC paradigm in one of two major ways. First, in some paradigms the nature of the CS–US relation changes dynamically over the course of the experiment. Such changes can be induced by introducing additional phases before or after the acquisition phase in which either the CS or the US is presented in isolation. These types of designs include latent inhibition (i.e., unpaired CS presentations before the CS–US pairing), extinction (i.e., unpaired CS presentation after the CS–US pairings), US pre-exposure (i.e., unpaired US presentations before the CS–US pairings), and US postexposure (i.e., unpaired US presentations after the CS–US pairings). Dynamical changes can also be induced by reversing the valence of the US with which the CS has been paired, such as in counterconditioning (i.e., an already conditioned CS is conditioned again with a US of opposite valence) or in US revaluation (i.e., the valence of the US is reversed after the CS and the US have been paired). Second, in some paradigms the relation between a CS (denoted here as CS1) and the US is (a) only indirect via another CS (CS2) that is or has been paired directly with the US or (b) context dependent. These designs include second-order conditioning (i.e., US is first paired with CS2; CS1 is then paired with CS2; change in liking of CS1 is of interest), sensory preconditioning (i.e., CS1 is first paired with CS2; CS2 is then paired with US; change in liking of CS1 is of interest), and occasion setting (i.e., discrete or context-stimulus *X* predicts whether CS and US co-occur).

Table 4 provides information about the mean effect size estimates with regard to all special designs for which at least three independent study effect size estimates were available: latent inhibition, extinction, US pre-exposure, second-order conditioning, and sensory preconditioning. It should be noted, however, that the number of studies for most of these paradigms was very small and, consequently, confidence intervals are very large (see Table 4). Hence, more research is needed before firm conclusions should be drawn. As can be seen from Table 4, all of these special designs produced mean effect size estimates that were significantly different from zero, except latent inhibition. Latent inhibition showed the smallest average effect, whereas US pre-exposure yielded the largest effect sizes in descriptive terms. However, because of the small number of available studies and the large confidence intervals, these differences between designs are far from statistical significance.[9]

The only special design applied in a substantial number of studies ($n = 20$) in our database was extinction. The average EC effect (d) after an extinction procedure (postextinction) was .558. A contrast test showed that this effect was not different in magnitude from the average (postacquisition) EC effect in the standard paradigm, that is, without an extinction procedure being present, $Q_B(1) = .04$, $p = .82$. Hence, from this perspective, EC appeared to be resistant to extinction. However, to scrutinize this conclusion further, we conducted a more fine-grained follow-up comparison in which we included only those 19 extinction studies for which both postacquisition and postextinction effect size estimates were available for the same outcome measure and for the same type of effect size contrast (i.e., in terms of LD/LN/DN contrast, and in terms of pre–post/post–post comparison). This follow-up analysis revealed that the above conclusion was premature: As can be seen from the bottom-most analysis presented in Table 4, the subset of extinction studies seems to have used paradigms that yield postacquisition EC effects of above-average magnitude ($d = .85$). Compared with this postacquisition effect, however, the postextinction effect ($d = .53$) was substantially reduced (by a magnitude of 37%), and the difference was significant (see Table 4). Taken together, the findings from this analysis point to the conclusion that even though EC is still present at postextinction, it does not appear to be resistant to extinction in the strict sense of the word.

Sensitivity Analysis

In meta-analysis, it is considered best practice to conduct a sensitivity analysis of the presented model (e.g., Borenstein et al., 2009; Greenhouse & Iyengar, 1994; Nelson & Kennedy, 2009). A sensitivity analysis addresses the question of whether results are affected substantially by variations in critical aspects of the criteria applied for selecting the main sample of analysis. To evaluate the robustness of our analyses, we conducted a sensitivity analysis on what we considered the four most important variations: In comparison with Model 1 ($n = 214$), we did not exclude outliers in Model 2 ($n = 215$ studies). In comparison with Model 1, we included only within-subjects data in Model 3 ($n = 166$). In comparison with Model 1, we included only effect sizes from LN and DN contrasts in Model 4 ($n = 151$). Finally, in Model 5 ($n = 253$), we included all dependent outcome measures in order to examine whether conclusions also hold across different valence assessment methods.[10]

For the overall analyses on these models, similar conclusions regarding the average effect size and its heterogeneity could be drawn (Model 2: $d = .53$, $SE = .03$, $I^2 = 70\%$; Model 3: $d = .56$,

[8] As expected from the way the formula was constructed, the contingency index was largely uncorrelated with the number of paired trials ($r = .07$, $p = .31$). Separate follow-up analyses including these two continuous predictors in a simultaneous regression analysis consequently yielded highly similar regression coefficients to those obtained in the single predictor analyses (across all analysis Models 1 to 5) and did not alter any of the conclusions drawn.

[9] An additional ANOVA in which the standard EC paradigm was included as an additional category yielded a nonsignificant overall effect, indicating that none of the special designs differed significantly from the standard EC effect, $Q_B(5) = 1.26$, $p = .939$.

[10] Separate sets of moderator analyses including only choice, implicit measures, or startle response data were not conducted because of the small number of studies for each of these categories.

SE = .03, I^2 = 73%; Model 4: *d* = .45, *SE* = .04, I^2 = 67%; Model 5: *d* = .48, *SE* = .02, I^2 = 63%). With regard to the moderator analyses, Table 5 provides an overview of the sensitivity analysis by giving the key statistics for each categorical or continuous moderator analysis separately for each type of analysis. As can be seen from this overview, statistical conclusions for moderator effects were highly identical across types of analyses. There were only five exceptions (see Table 5): First, the moderator effect of CS Modality was no longer significant when considering only within-data or all dependent outcome measures. Hence, the above-described difference between verbal sensical and verbal nonsensical material should be interpreted with caution. Second, third, and fourth, respectively, the effects of sample, US selection, and CS test showed just a trend toward significance when we included only within-subjects data in the analysis. Fifth, the number of paired trials was significantly positively related to effect sizes for the within-subjects data analysis, a finding that is further discussed below.

In order to quantitatively describe the degree of convergence of estimates, we calculated the absolute difference between the category means for the main model (i.e., the *d* estimates reported in Tables 3 and 4) and each corresponding category mean for the four alternative models. The mean absolute difference in *d* estimates was Δd = .001 for the no-outlier-exclusion model, Δd = .028 for the within-data only model, Δd = .051 for the model including only LN and DN contrasts, and Δd = .049 for the model including all outcome measures. An analogous comparison of the corresponding *SE* estimates between Model 1 and the other models yielded a close fit as well (ΔSEs = .001, .007, .011, and .007 for Models 2, 3, 4, and 5, respectively). Taken together, the present meta-analytic conclusions and estimates appear to be quite robust with regard to alternative strategies of analysis.

Additional Findings Involving All Dependent Outcome Measures (Model 5): Type of Dependent Variable, Implicit Measure, and Supraliminal/Subliminal Conditioned Stimulus

The inclusion of all dependent outcome measures in Model 5 enabled a few additional analyses of interest. First, we investigated whether EC effects differ with regard to the type of dependent outcome measure used in order to assess changes in the valence of the CS. Corroborating the preliminary analysis above, we found that type of dependent variable accounted for significant amounts of variance in EC effects (see Table 5, right column). As indicated by contrasts, EC effects assessed with implicit measures of valence (*M* = .298, *SE* = .050, *K* = 57) were significantly smaller in magnitude than those for self-reports (*M* = .529, *SE* = .026, *K* = 241), choice measures (*M* = .543, *SE* = .102, *K* = 20), or the physiological measure of startle response magnitude (*M* = .505, *SE* = .142, *K* = 9), the latter three of which did not differ reliably from each other. A follow-up analysis on the subsample of implicit measures (see Table 5, rightmost columns) revealed significant variability in effect sizes. Specifically, EC effects, as assessed with the affective priming paradigm (*M* = .200, *SE* = .041, *K* = 30), albeit significantly greater than zero, were significantly smaller than EC effects assessed with the Implicit Association Test (*M* = .396, *SE* = .060, *K* = 21) and EC effects assessed with the Name Letter Task (*M* = .507, *SE* = .152, *K* = 3), with no significant difference between the latter two. Finally, Model 5 allowed for a tentative test of the contrast between CS subliminal and supraliminal presentation (see Table 5, rightmost columns), as most of the CS subliminal studies in our data set involved implicit measures of valence assessment. However, the number of study effect sizes including subliminal CS presentation was still relatively low (*n* = 8), and estimated effects for subliminal (*M* = .490, *SE* = .137, *K* = 8) and supraliminal (*M* = .460, *SE* = .027, *K* = 201) presentation were not significantly different from each other (see Table 5, rightmost columns).

Relationships Among Moderators

Although a simultaneous inclusion of multiple moderator variables in one and the same analysis (i.e., by multiple regression with dummy coded and continuous variables) would be desirable in order to reduce potential redundancies or confounds, a simultaneous test was not feasible on the level of study effect sizes. This was the case because, as a result of the variation of a subset of moderator variables within studies, the aggregated study effect sizes (i.e., the dependent variable) changed depending on which moderator variable was investigated. A complete simultaneous analysis would have been possible only on the level of single effect sizes, but this clearly would have violated the assumption of independence. Also, a simultaneous analysis would have greatly reduced the power of the analysis because of missing values in cases where a definite coding could not be made and because of the extremely high number of predictors in the regression model resulting from simultaneous inclusion and from the dummy coding of categorical moderators. In order to get a sense of whether there were any strong redundancies or confounds among the moderators investigated in the present analysis, we checked the bivariate correlations among all moderators in the data set for the full analysis (all dependent variables) on the level of effect size codings. As coefficients, we computed (a) the absolute value of Pearson's *r* for relations among continuous moderators, (b) Cramer's *V* for relations among dichotomous/polytomous categorical moderators, and (c) the multiple (unsquared) correlation coefficient *R* for the relationship among continuous moderators and dichotomous/polytomous categorical moderators. For all of these indices, a value of zero indicates complete independence, whereas a value of 1 indicates a perfect relationship. On average, relationships among moderators were very low (mean *r* = .12; mean *V* = .21; mean *R* = .14).[11] When considering as substantial only those relationships whose absolute magnitude exceeded a value of .50, only a small number of 10 substantially related pairs of moderators emerged (3.9% of all coefficients). Seven of these substantial relationships could be directly traced to the close conceptual or procedural relatedness among moderators: *r*(CS duration/US duration) = .99; *r*(CS-only/contingency index) = −.63; *r*(US-only/contingency index) = −.67; *r*(interstimulus interval/intertrial interval) = .96; *R*(contingency awareness, aware %) = .61; *R*(contingency/CS-only) = .78; and *R*(contingency/contingency index) = .81. The remaining three substantial correlations represent "true" confounds whose nature and effects were investigated

[11] A table with all intercorrelations can be obtained from Wilhelm Hofmann on request.

Table 5
Overview of Sensitivity Analysis

	1. Main model				2. No outlier exclusion				3. Within-subjects data only				4. LN and DN only				5. All dependent variables			
Moderator	Q_B/β	$df_{(B,W)}$	*p*	Level	Q_B/β	$df_{(B,W)}$	*p*	Level	Q_B/β	$df_{(B,W)}$	*p*	Level	Q_B/β	$df_{(B,W)}$	*p*	Level	Q_B/β	$df_{(B,W)}$	*p*	Level
Sample	16.01	(5, 208)	.01	**	15.90	(5, 209)	.01	**	10.79	(5, 160)	.06		16.11	(5, 145)	.01	**	11.07	(5, 247)	.05	*
Gender (β)	0.01	(1, 115)	.89		0.01	(1, 116)	.87		0.05	(1, 106)	.61		0.02	(1, 80)	.82		−0.07	(1, 134)	.36	
CS modality	12.55	(5, 209)	.03	*	12.55	(5, 209)	.03	*	9.07	(5, 160)	.11		11.79	(5, 146)	.04	*	5.55	(6, 247)	.48	
US modality	48.31	(7, 211)	.00	**	45.99	(7, 212)	.00	**	48.31	(7, 162)	.00	**	56.59	(7, 146)	.00	**	25.39	(7, 249)	.00	**
CS–US modality match	0.33	(1, 146)	.56		0.61	(1, 149)	.43		2.57	(1, 104)	.11		0.04	(1, 117)	.84		0.22	(1, 163)	.64	
CS selection	15.11	(2, 188)	.00	**	15.11	(2, 188)	.00	**	7.66	(2, 150)	.02	*	7.60	(2, 132)	.02	*	9.75	(3, 219)	.02	*
US selection	8.55	(2, 211)	.01	*	7.44	(2, 212)	.02	*	5.27	(2, 163)	.07		9.12	(2, 149)	.01	*	9.00	(2, 248)	.01	*
CS duration (β)	0.07	(1, 190)	.29		0.07	(1, 191)	.31		0.07	(1, 146)	.34		0.03	(1, 133)	.72		0.08	(1, 220)	.21	
US duration (β)	0.08	(1, 183)	.23		0.08	(1, 184)	.24		0.08	(1, 138)	.29		0.04	(1, 138)	.62		0.09	(1, 208)	.15	
CS supra-/subliminal																	0.05	(1, 207)	.83	
US supra-/subliminal	8.01	(1, 176)	.00	**	8.15	(1, 177)	.00	**	6.38	(1, 133)	.01	*	6.40	(1, 131)	.01	*	7.02	(1, 201)	.01	**
A priori CS–US Match	0.13	(1, 179)	.72		0.15	(1, 180)	.70		0.07	(1, 160)	.79		0.00	(1, 118)	.99		0.01	(1, 212)	.94	
Same assignment	0.09	(1, 206)	.76		0.03	(1, 207)	.87		1.96	(1, 158)	.16		0.34	(1, 146)	.56		2.11	(1, 241)	.15	
Type of dependent variable																	17.09	(3, 323)	.00	**
Implicit measure																	9.73	(2, 51)	.01	**
CS test	19.43	(1, 213)	.00	**	19.28	(1, 214)	.00	**	3.00	(1, 165)	.08		15.80	(1, 149)	.00	**	14.41	(1, 252)	.00	**
Awareness	68.98	(3, 96)	.00	**	64.92	(3, 97)	.00	**	74.54	(3, 71)	.00	**	6.92	(1, 61)	.01	**	46.75	(3, 107)	.00	**
% Aware (β)	0.37	(1, 83)	.00	**	0.33	(1, 84)	.00	**	0.39	(1, 66)	.00	**	0.47	(1, 55)	.00	**	0.38	(1, 96)	.00	**
Learning	1.67	(1, 198)	.20		1.67	(1, 198)	.20		3.14	(2, 163)	.21		1.12	(1, 137)	.29		0.16	(1, 232)	.69	
Spontaneity	7.79	(1, 207)	.01	**	7.96	(1, 208)	.00	**	10.13	(1, 160)	.00	**	4.05	(1, 146)	.04	*	13.16	(1, 247)	.00	**
Contingency	0.01	(1, 212)	.94		0.01	(1, 213)	.91		0.54	(1, 164)	.46		0.35	(1, 148)	.56		0.03	(1, 250)	.86	
No. CS only (β)	0.00	(1, 215)	.95		0.00	(1, 216)	.98		−0.01	(1, 169)	.93		0.05	(1, 148)	.51		−0.09	(1, 261)	.11	
No. US only (β)	−0.07	(1, 218)	.24		−0.07	(1, 219)	.24		−0.09	(1, 170)	.19		0.00	(1, 151)	.98		−0.06	(1, 263)	.28	
Contingency index (β)	0.05	(1, 201)	.44		0.05	(1, 202)	.42		0.07	(1, 158)	.37		−0.03	(1, 137)	.69		0.09	(1, 240)	.14	
No. paired trials (β)	0.05	(1, 211)	.40		0.05	(1, 212)	.40		0.21	(1, 160)	.00	**	−0.04	(1, 148)	.60		0.09	(1, 253)	.13	
Presentation	1.10	(2, 200)	.58		0.86	(2, 201)	.65		0.77	(2, 154)	.68		0.21	(2, 138)	.90		0.37	(2, 237)	.83	
ISI (β)	−0.01	(1, 177)	.88		−0.01	(1, 178)	.86		−0.02	(1, 140)	.79		−0.03	(1, 130)	.71		0.01	(1, 209)	.92	
ITI (β)	0.00	(1, 155)	.96		0.01	(1, 156)	.87		−0.02	(1, 118)	.82		0.00	(1, 115)	.97		0.02	(1, 176)	.82	
Special designs	2.44	(4, 29)	.66		2.44	(4, 29)	.66		2.29	(4, 34)	.68		2.02	(3, 26)	.57		2.29	(4, 34)	.68	
Extinction	4.82	(1, 36)	.03	*	4.82	(1, 36)	.03	*	5.21	(1, 36)	.02	*	4.61	(1, 34)	.03	*	4.35	(1, 44)	.04	*

Note. Models 1 to 4 involve only self-report data; Model 5 includes all dependent outcome measures. Key statistics from the main model (Model 1) reported in Tables 3 and 4 are reproduced for comparison purposes. Empty cells indicate analyses for which a moderator analysis was not feasible (requirement: at least two moderator categories with a minimum of three study effect sizes each). LN = liking-neutral comparison; DN = disliking-neutral comparison; CS = conditioned stimulus; US = unconditioned stimulus; ISI = interstimulus interval; ITI = intertrial interval; Q_B = analysis of variance (ANOVA) between-groups sum of squares; β = standardized regression coefficient; $df_{(B,W)}$ = degrees of freedom for between- and within-groups analysis; *p* = significance level of between-groups effect (ANOVA) or regression coefficient.
* $p < .05$. ** $p < .01$. *p* values of .00 indicate numbers smaller than .005.

further. In all three cases, CS modality was involved. First, CS modality was substantially related to US modality ($V = .55$), reflecting the fact that a majority of studies using USs of a visual (87%), taste/flavor (85%), or haptic modality (100%) also used the same modality for the CS. Therefore, because of these choices made on the level of the primary research, the significant US and CS modality effects of the present analysis cannot be regarded as completely independent of each other. Second, CS modality was substantially related to CS test ($V = .63$). This relation was due to the fact that 69% of studies involving a different CS than the one presented during the acquisition phase also used sensical verbal material as CS, whereas the remaining 31% used visual material. Hence, the smaller effect obtained for different CSs in our analysis may be (at least partly) due to the fact that it is more difficult to imbue sensical verbal material with new evaluative meaning than with new visual material. Third, CS modality and subliminal CS were substantially related ($V = .74$), reflecting the fact that 100% of the studies involving subliminal CS presentation used sensical verbal material, whereas more variable combinations of CS modality with supraliminal presentation were used. As pointed out previously, the CS subliminal moderator involved a very low number of primary studies. Hence, more research is needed in any case irrespective of this potential confound.

Taken together, then, the lion's share of moderator analyses in our meta-analysis reflects cases for which substantial confounds with other moderators in our data set can be ruled out. However, there are a few cases, as mentioned above, where more diverse primary research is needed before stronger conclusions about which factor seems to be the driving force behind observed variations in EC effects can be drawn.

Discussion

How humans acquire their likes and dislikes can be illuminated by the study of EC, that is, the way in which the pairing of stimuli alters the liking of those stimuli. In the present article, we report results from a large-scale meta-analysis of evaluative conditioning research. Our quantitative summary was informed by three main questions: (a) Is EC a genuine and general phenomenon? (b) Is EC a unique form of Pavlovian conditioning? (c) What are the processes underlying EC? In the remainder of this article, we highlight what we believe are the conclusions from our meta-analysis regarding these three questions and place the present meta-analytic evidence in the broader context of ongoing debates in the EC literature.

Is Evaluative Conditioning a Genuine and General Phenomenon?

The first part of this question ("Is EC a genuine phenomenon?") can be answered from our global assessment of the magnitude of EC effects. Across our main sample of 214 primary studies, the mean EC effect was close to what is considered a medium effect size (J. Cohen, 1977). From this global assessment, we can conclude beyond any doubt that EC is a genuine phenomenon. The second part of the question ("Is EC a general phenomenon?") can be answered from the assessment of the degree of heterogeneity in effect sizes. Our random effects estimates indicated that more than two thirds of the variance in effect sizes across studies can be attributed to systematic sources rather than to sampling error. Hence, even though EC exists as an authentic, genuine phenomenon, EC effects do not always occur. Rather, the substantial degree of heterogeneity suggests that in order to truly understand EC, it is important to elucidate the boundary conditions responsible for the systematic variation in EC effects across studies. The question, therefore, is not so much whether EC exists but rather *when* it is expected to lead to strong as opposed to weak changes in preferences. Identifying the key variables that are able to explain significant portions of variation in EC effects can inform us both about the specific boundaries of EC, its similarity or dissimilarity from other forms of conditioning, and about the theoretical processes underlying EC.

Is Evaluative Conditioning a Unique Form of Pavlovian Conditioning?

The answer to this second question depends on whether there are potential moderators that have a different impact on EC than on other forms of Pavlovian conditioning. We believe that our evidence challenges some previous claims about the ways in which EC differs from other forms of conditioning. First, contrary to the claim that EC does not depend on contingency awareness, the present results revealed that contingency awareness was by far the most important moderator of EC, accounting for as much as 36% of the variance of the EC effects. Our analyses showed that (a) participants who are classified as contingency aware show a larger EC effect than participants who are classified as contingency unaware, (b) EC is stronger when considering only CSs for which participants were contingency aware than when considering only CSs for which participants were contingency unaware, and (c) EC effects were larger in studies with many contingency-aware participants than in studies with few contingency-aware participants (% aware). Across the seven studies in which contingency awareness was assessed on an item-to-item basis, EC was not significant when considering only CSs for which participants were contingency unaware. Additional evidence stems from the abovementioned finding that subliminal US presentations did not produce significant EC effects. The only finding that supports the idea of unaware EC was the presence of a strongly reduced but still significant EC effect across 48 samples of contingency-unaware participants. However, this finding must be interpreted with caution because of the potentially limited validity of the way in which participants were divided into contingency-aware and contingency-unaware groups (see Field, 2000; Lovibond & Shanks, 2002; Pleyers et al., 2007, for more details). For a long time, it has been argued that EC differs from other types of Pavlovian conditioning in that it is independent of contingency awareness (e.g., Baeyens & De Houwer, 1995; Martin & Levey, 1978). Our meta-analysis shows that this position is no longer tenable. Our findings leave little doubt that contingency awareness is an important moderator of EC just as in Pavlovian conditioning (e.g., Lovibond & Shanks, 2002; Mitchell, De Houwer, & Lovibond, 2009b). The meta-analysis thus contradicts the hypothesis that EC is unique in that it is independent of contingency awareness.

Second, the present results also raise doubts about the claim that, unlike other forms of Pavlovian conditioning, EC is resistant to extinction. Fine-grained analyses showed that unpaired CS presentations after the CS–US trials (i.e., extinction) reduce the magnitude of EC. Hence, EC is sensitive to extinction. This is an important conclusion that clearly demonstrates the benefits of a meta-analysis. Whereas many individual, low-powered studies failed to provide evidence for extinction, extinction is found when the available data are aggregated across studies. At the same time, however, our analysis also showed that a substantial EC effect remained even after unpaired CS presentations. Although more studies are needed that directly compare the rate of extinction in EC and other forms of Pavlovian conditioning, it is possible that extinction occurs at a slower rate in EC than in other forms of Pavlovian conditioning (e.g., Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006).

The results of our meta-analysis are compatible with a third claim about the uniqueness of EC, however. They suggest that the degree of statistical contingency between the CS and the US during acquisition has little, if any, effect on the magnitude of EC. The fact that other forms of Pavlovian conditioning strongly depend on the statistical contingency between the CS and the US (e.g., Rescorla, 1966) suggests that EC is unique in this respect. Hence, it is possible that EC may be determined by other aspects of the CS–US relation than other forms of Pavlovian conditioning. More specifically, whereas Pavlovian conditioning is mainly determined by the strength of the statistical relation between the CS and US, there was some (limited) indication that EC seems to be more strongly influenced by the number of times that the CS and the US have co-occurred. Note that this conclusion concerns just the effect of CS-only and US-only presentations that are intermixed with the CS–US trials during acquisition. Our analyses showed that CS-only trials that are presented before (i.e., latent inhibition) or after the CS–US trials (i.e., extinction) do reduce the magnitude of EC. Thus, the impact of CS-only trials on EC seems to depend on the time at which the CSs are presented (before, during, or after the CS–US trials).

In sum, the results of our meta-analysis show that EC is less unique than is often assumed. Like other forms of Pavlovian conditioning, EC depends heavily on contingency awareness and is sensitive to extinction. It does, however, seem to be less influenced by the statistical contingency between the CS and the US than are other forms of Pavlovian conditioning.

What Are the Processes Underlying Evaluative Conditioning?

The answer to the third and final question is determined by whether theories of EC can explain why EC is affected by certain but not other moderators. In the following passages, we discuss the main moderator findings in light of the five theoretical accounts of EC introduced earlier. Unless further differentiation is warranted, we consider the referential, holistic, and misattribution account together under the umbrella of association formation models. Moreover, in discussing the compatibility between findings and accounts, we distinguish between cases where a given finding can be inferred directly from the central assumptions of a given account (for an overview, see Table 1) and cases in which a given finding can be reconciled with an account if an additional auxiliary assumption is introduced (specified in the text). An overview of our conclusions is given in Table 6.

1. Evaluative conditioning effects are smaller in children. EC appears to be largely independent of the nature of the sample in which it is studied. Effects in children, however, appear to be considerably smaller than those observed in adults (see also O'Donnell & Brown, 1973). Assuming that association formation is a largely automatic process that should be fully functional by an early age, this finding is puzzling from the perspective of association formation models of EC, such as the referential, the holistic, and the implicit misattribution accounts. It does, however, fit with the idea that EC is based on the nonautomatic formation and evaluation of propositions about CS–US relations. Assuming that children are poorer at consciously identifying CS–US relations, the propositional approach can explain that EC effects tend to be smaller in children. Further, the conceptual categorization account can accommodate this finding if it is assumed that the conceptual learning mechanisms presumed to underlie EC represent a higher order mental process that is not yet fully developed in children. It should be noted, however, that only a small number of EC studies involved children as participants. The conclusion that EC is smaller in children should thus be interpreted cautiously.

2. Evaluative conditioning effects are larger (a) for nonsensical as compared with sensical verbal conditioned stimuli and (b) for neutral as compared with initially valenced conditioned stimuli. There was some indication that EC is smaller for sensical verbal CSs as compared with nonsensical CSs (CS modality). EC effects were also stronger when pretesting ensured that the CS was affectively neutral as compared with initially valenced CSs. Viewed in concert, these two results seem to speak to the issue of attitude formation in comparison with attitude change. Specifically, a nonsensical CS may be relevant only to attitude formation and not to attitude change because, by definition, one cannot have an attitude toward nonexisting categories. In a similar vein, an evaluatively neutral CS often may imply that no existing affective attitude has yet been formed. Though somewhat speculative, EC effects may thus be stronger for the formation of new attitudes than for the change of existing attitudes.

Both association formation and propositional models can provide an explanation for this pattern of results. From an association formation perspective, it has been argued that forming a new association may be easier than changing a pre-existing association (e.g., Gregg, Seibt, & Banaji, 2006). The implicit misattribution account, in particular, adds to this explanation by postulating that an association between a CS and a US should be easier to form when the CS has a high degree of ambiguity—such as a CS of a nonsensical or affectively neutral nature. A propositional explanation is also possible if one assumes that propositions about CS–US relations have more relative impact on judgments about the valence of the CS if the liking of the CS is not already based on a meaningful set of pre-existing propositions. Because propositions are assumed to be subject to mechanisms of cognitive consistency (e.g., Gawronski & Bodenhausen, 2006), adding new propositions to an already existing set may often result in less attitude change than creating a new proposition on the spot. The conceptual categorization account, in contrast, may have some difficulties in explaining Finding 2a because it is implausible to assume that nonsensical verbal material contains enough salient features that

Table 6
Overview of Main Moderator Findings and Their Relation to Theoretical Accounts of Evaluative Conditioning

	Relation to major theoretical accounts of evaluative conditioning				
Main moderator findings	Referential	Holistic	Misattribution	Categorization	Propositional
1. EC effects smaller in children	Incompatible	Incompatible	Incompatible	Compatible (AA)	Compatible (AA)
2a. EC effects larger for nonsensical verbal CSs as compared with sensical CSs	Compatible	Compatible	Compatible	Incompatible	Compatible
2b. EC effects larger for neutral CSs as compared with initially valenced CSs	Compatible	Compatible	Compatible	Compatible	Compatible
3. EC effects larger when electrocutaneous stimulations is used as US	Compatible (AA)	Compatible (AA)	Incompatible	Incompatible	Compatible (AA)
4a. EC independent of whether CS and US are matched in terms of stimulus modality	Compatible (AA)	Compatible (AA)	Incompatible	Incompatible	Compatible
4b. EC independent of whether CS and US are matched in terms of perceptual similarity	Compatible (AA)	Compatible (AA)	Incompatible	Incompatible	Compatible
5. EC effects larger for supraliminal than subliminal US presentations	Incompatible	Incompatible			Compatible
6. EC effects smaller with regard to implicit measures of liking than self-report measures	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
7. Significant implicit measure effect/startle response effect	Compatible	Compatible	Compatible		Incompatible
8a. EC effects larger for contingency awareness in subsamples/on level of pairings	Incompatible	Incompatible	Incompatible		Compatible
8b. EC effects present for contingency unaware participants	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
9. Spontaneous evaluation of CS results in reduced EC effect	Incompatible	Incompatible	Incompatible	Incompatible	Compatible
10a. No effect of statistical contingency	Compatible	Compatible	Compatible	Compatible	Compatible (AA)
10b. Some evidence for CS–US co-occurrence (no. paired trials) increasing EC effects (Analysis 3)	Compatible	Compatible	Compatible	Compatible	Compatible (AA)
11. EC sensitive to extinction when comparing post-acquisition and post-extinction effects	Compatible (AA)	Compatible (AA)		Compatible (AA)	Compatible

Note. Moderator findings are presented in order of moderator framework. Compatible/incompatible = finding is supported/not supported by the core assumptions of the theoretical account in question; AA = finding can be reconciled by introducing an auxiliary assumption (see text for details); inconclusive: finding does not allow for a clear interpretation because of a serious methodological caveat (see text for details). Empty cells represent cases where the theoretical account does not appear to provide enough specification.

can be highlighted through the pairing with a liked or disliked US. This account can, however, easily explain Finding 2b because neutral material may provide a better starting point for selectively highlighting positive or negative features of the CS through pairing with liked or disliked stimuli.

3. Evaluative conditioning effects are larger for electrocutaneous stimuli. EC appears to be stronger when electrocutaneous stimulation is used as the US than when other types of USs are used. An obvious step in interpreting this finding is to assume that electrocutaneous stimuli are more evaluatively intense—negative, to be precise—than most other USs used. The referential account and the holistic account can accommodate such a US intensity explanation by assuming that the formation of associations is (biologically) facilitated for intense evaluative US responses or that presentation of the CS leads to the activation of more intense evaluative US responses. On the contrary, the misattribution account predicts that salient affective experiences should be correctly attributed to their true source (the US) and, hence, a misattribution of these experiences to the CS should become less likely. From the perspective of the conceptual categorization account, it is again difficult to think of the types of salient features that receiving a shock selectively highlights in other types of stimulus modalities. From a propositional stance, this finding could be due to the fact that participants are more motivated to detect relations involving highly salient and relevant stimuli, such as electrocutaneous stimulation.

4. Evaluative conditioning is independent of conditioned stimulus–unconditioned stimulus match. EC appears to be independent of whether CS and US are matched (a) in terms of stimulus modality (CS–US modality match) or (b) according to certain criteria such as perceptual similarity (a priori CS–US match). Taken together, these two findings suggest that EC does not appear to be sensitive to specific stimulus constellations. The observation that CS–US modality match did not influence EC seems to be incompatible with the implicit misattribution account and the conceptual categorization account. Given that source confusion and conceptual categorization mechanism can operate more strongly when the CS and the US have features in common, both accounts would predict that EC occurs more strongly when the CS and the US are of the same modality than when they are of different modalities. The fact that CS–US modality match and CS–US match did not influence EC speaks to the generality with which contingencies can become mentally represented, irrespective of any a priori matches or mismatches between stimuli. These findings are compatible with the referential and holistic account as long as it is assumed that the automatic link–formation mechanism is not facilitated or impeded with regard to certain CS–US constellations. They are also compatible with a propositional account because, in principle, any kind of CS–US relation can be mentally represented in a propositional format.

5. Evaluative conditioning effects are larger for supraliminal unconditioned stimulus presentations. EC effects were markedly larger for supraliminal than for subliminal US presentations, and the latter effect was not significantly different from zero. Because subliminal presentation prevents participants from becoming aware of the CS–US relation, this finding fits well with the effect of contingency awareness and can be accounted for by propositional rather than associative accounts of EC (see discussion below). However, analyses also showed that the effect of CS subliminal presentations did not differ from the effect of supraliminal stimulus presentations. It should be noted that the latter analyses involved only a very small number of studies (stemming from only three different citations) and was primarily composed of implicit measures. Whether this nonsignificant moderator effect is due to low power, differences in measurement, or differences in the degree to which the US rather than the CS needs to be consciously represented for EC to occur is an attractive avenue for future research because, to date, no study has investigated the effects of subliminal and supraliminal US/CS presentations in a fully crossed design.

6. Evaluative conditioning effects are smaller when implicit measures of liking are used. Reliable EC effects were obtained with a variety of indices of liking. Still, we discovered some interesting differences between different dependent outcome measures. The observation that EC effects were smaller with implicit measures of liking than with other measures at first sight goes against association formation models of EC. From the background of these models, one would assume that implicit measures should reflect the nature of CS–US associations in memory in a relatively direct manner (e.g., Hermans, Baeyens, Lamote, Spruyt, & Eelen, 2005). Implicit measures such as affective priming should therefore be at least equally potent indicators of EC as self-report measures of liking. In contrast to this view, the larger empirical effect in self-report measures seems to favor a propositional explanation. The gradual decline of EC effects from self-report to Implicit Association Test (IAT) scores to affective priming measures also seems consistent with a propositional interpretation if one assumes that differences between IAT and affective priming effects are due to a higher sensitivity of the IAT with regard to propositional influences (e.g., De Houwer, 2006). However, there is a psychometric caveat here that precludes drawing firm theoretical conclusions—the reliability of measurement. Specifically, affective priming measures have been repeatedly criticized for their low reliability (e.g., Cunningham, Preacher, & Banaji, 2001), and these differences may account for large portions of this moderator effect. Note, however, that an interpretation in terms of reliability faces some difficulties in explaining the gap between the magnitude of EC effects for self-report as compared with IAT measures because internal consistencies of the IAT typically are around .80 (e.g., Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005) and thus similar to those commonly obtained for self-reports.

7. Significant implicit measure/startle effects. Even though effect sizes for implicit measures of liking, such as affective priming, were smaller than in self-report data, these effects were still significantly greater than zero. In addition, EC was manifest in a medium-sized average startle response effect. EC effects in implicit and physiological measures can be accounted for by association formation models as a conditioned response that results from the automatic associative activation of the US representation. The propositional account can only explain EC effects in implicit and physiological measures if it is additionally assumed that these indirect measures are not impervious to higher order cognitive processes (see De Houwer, 2006; De Houwer et al., 2005). It is true, however, that the propositional account in its present form is relatively mute with regard to the possible interplay of propositional and associative representations (see Gawronski & Bodenhausen, 2006; Hofmann, Gschwendner, Nosek, & Schmitt, 2005, for models specifying such an interplay) and with regard to the translation of propositional beliefs into specific physiological responses such as eyeblink startle reflexes (Baeyens, Vansteenwegen, & Hermans, 2009; but see Mitchell, De Houwer, & Lovibond, 2009a, for a response to this criticism).

8. Evaluative conditioning effects are (a) larger for contingency awareness in subsamples/on the level of pairings and (b) significantly different from zero in participants classified as contingency unaware. As mentioned above, contingency awareness was by far the most potent moderator of EC, and this moderator effect was robust with regard to a number of analysis strategies (such as interindividual and intraindividual assessments of contingency awareness and subliminal vs. supraliminal US presentations). Taken together, there thus can be little doubt that contingency awareness is an important moderator of EC. The observation that contingency awareness is such a strong moderator of EC cannot be accounted for by existing associative models of EC such as the referential account or the holistic account. According to these models, CS–US associations are formed and influence liking in an automatic manner, that is, regardless of whether participants are aware of the CS–US contingency. From the perspective of the misattribution account, the observed positive relation between contingency awareness and EC effects even goes in the opposite direction of what would be expected if one assumes that contingency awareness may offset the misattribution mechanism (Jones et al., 2009). The strong impact that contingency awareness has on EC is, however, entirely in line with a propositional account of EC. According to this account, participants need to form a conscious proposition about the CS–US relation before this relation can influence liking of the CS (De Houwer, 2007a; De Houwer et al., 2005). Hence, EC should depend on whether participants are aware of the CS–US relation. However, the propositional account would be unable to explain EC in the absence of contingency awareness, whereas association formation models would be compatible with such effects.

Strongly reduced but still significant EC was found, however, across 48 samples of contingency-unaware participants. From a theoretical stance, this finding is compatible with associative models and incompatible with the propositional account. It may indicate that automatic associative processes produce at least very weak EC effects in the absence of contingency awareness. From a methodological viewpoint, however, strong doubts have been raised about the validity of the way in which participants were divided into contingency-aware and contingency-unaware groups (see Field, 2000; Lovibond & Shanks, 2002; Pleyers et al., 2007, for more details). The reduced sensitivity of a participant-wise classification may also explain why recent studies conceptualizing awareness in a more fine-grained manner at the (intraindividual) level of pairings (e.g., Pleyers et al., 2007; Stahl & Unkelbach, 2009) found a more pronounced aware–unaware discrepancy (with pairings classified as unaware yielding no significant EC effects). Hence, from the perspective of the present meta-analysis, it appears unwarranted to draw firm theoretical conclusions about the finding of significant EC in participant samples classified as contingency unaware.

Taken together, the results of the present meta-analysis yield strong evidence for contingency awareness as a key factor in EC and relatively weak evidence for EC in the absence of contingency awareness. Even though EC seems to be a potent moderator, it is unclear at present whether contingency awareness is a *necessary* condition for EC to occur and what, if any, causal role contingency awareness plays in mediating EC effects (e.g., Lovibond & Shanks, 2002). Future research will have to clarify these issues by creating innovative new paradigms with which contingency awareness can be manipulated experimentally rather than by relying on postacquisition memory measures of contingency awareness alone.

9. Spontaneous evaluation of conditioned stimuli results in reduced evaluative conditioning effect. The meta-analysis shows that the instruction to evaluate the CSs spontaneously weakens EC. This finding is also surprising from the perspective of association formation models of EC and contradicts suggestions in the literature that EC effects are more robust when participants are instructed to evaluate the stimuli in a spontaneous manner (e.g., De Houwer et al., 2005, p. 167). It also is not clear to us how this finding can be reconciled with the conceptual categorization account. The effect of this moderator is, although somewhat speculative, compatible with a propositional account of EC. If EC is based on propositions about CS–US relations, effects might be particularly strong when the circumstances allow participants sufficient processing time and/or encourage participants to intentionally use these propositions as a justification for evaluating the CS.

10. Evaluative conditioning (a) is independent of statistical contingency but (b) seems to depend primarily on conditioned stimulus–unconditioned stimulus co-occurrences. Our meta-analysis suggests that the degree of statistical contingency between the CS and the US has little if any effect on the magnitude of EC. There was some indication that EC becomes stronger when the number of CS–US co-occurrences (number of paired trials) increases, but note that this finding was limited to the analysis of within-participants designs (Model 3). This finding lends some support for association formation models of EC according to which EC should be determined mainly by the number of times that the CS and the US co-occur rather than by the number of times that they are presented in isolation. It is also consistent with the conceptual categorization account because a higher number of pairings should increase the salience of shared features between the CS and the US. This finding seems to be at odds, however, with the propositional account arguing that statistical contingency should foster the propositional belief that the CS and the US are related. It is possible, in principle, to reconcile these findings with the propositional account by assuming that EC does not always depend on the formation of propositions about the statistical contingency between the CS and the US but can also result from the formation of propositions about the co-occurrence of the CS and US. Such an explanation is clearly post hoc and calls for an empirical content analysis or experimental manipulations of participants' conscious beliefs about CS–US relations.

11. Extinction effect. Fine-grained analyses showed that unpaired CS presentations do reduce the magnitude of EC when they are presented after the CS–US trials. Hence, EC is sensitive to extinction. The notion that EC is, to some degree, sensitive to extinction goes against earlier claims of association formation models of EC (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988). These models have emphasized the role of CS–US co-occurrences as the main determinant of EC effects and the stability of associations in long-term memory once they have been formed. The fact that EC is, to some degree, sensitive to extinction is not predicted by these models but can be reconciled if it is additionally assumed that unpaired CS presentations after CS–US pairings produce habituation of the liking of the CS, rather than unlearning of the CS–US association, or produce the additional learning of a CS–no-US association (e.g., Bouton, 2004). The conceptual categorization account can also accommodate the extinction effect by assuming that CS-only trials reduce the salience of features shared with the US. From the perspective of the propositional account, the extinction effect can be explained by assuming that repeated CS-alone presentations may lead participants to consciously adjust their propositional belief that the CS and the US are related in a certain manner and that, as a consequence, the CS loses some of its previously acquired valence. Note, however, that this interpretation in terms of statistical contingency makes sense only in those cases where EC is based on propositions about the statistical contingency between CS and US.

Summary and conclusions regarding theoretical accounts of evaluative conditioning. Taken together, the results of the meta-analysis provide important information about the mental processes that underlie EC. Existing association formation models, such as the referential account and the holistic account, provide a parsimonious account for the greater-than-zero EC effects in affective priming and startle response data, and they correctly predict that EC is driven mainly by co-occurrences of the CS and the US. Yet, association formation models are unable to provide a straightforward explanation for the effect of several moderators of EC. Most important, they cannot account for the huge effect of contingency awareness on EC. In a related vein, they cannot explain the absence of a reliable mean EC effect with regard to subliminal US presentations. Association formation models also have difficulties accounting for the fact that children show smaller EC effects than adults and that receiving instructions to evaluate the CS spontaneously reduces the magnitude of EC.

It is possible, in principle, that new associative models that are more compatible with the results of our meta-analysis will be proposed in the future. For instance, rather than assuming that association formation is an automatic process, one could postulate that the formation of associations in memory or the impact of these associations on liking depends on awareness of the CS–US relation and thus on all factors that influence contingency awareness (e.g., Dawson & Schell, 1985). Although association formation models can be modified in this manner, there are no a priori reasons to assume that (the impact of) association formation should depend on contingency awareness. Moreover, adding the assumption that association formation depends on contingency awareness calls into question the widespread assumption that association formation is a basic learning mechanism that operates across species (see also De Houwer, 2009b; Mitchell et al., 2009a).

The recently proposed implicit misattribution account of EC shares some assumptions with the holistic account as an association formation account. However, the former account emphasizes the conditions under which a transfer (i.e., misattribution) of valence from the US to the CS is most likely. Together with the conceptual categorization account, this account postulates a key role for feature overlap between the CS and the US as a moderator of EC. This prediction, however, is unsupported by the present finding that EC effects are relatively unaffected by variables related to feature overlap (i.e., a priori CS–US match, CS–US modality match). This does not imply that implicit misattribution or conceptual categorization may not operate in EC at all (for empirical support, see Jones et al., 2009). However, it is questionable whether implicit misattribution or conceptual categorization as explanatory mechanisms are by themselves strong and general enough to account for the full range of the present meta-analytic evidence.

The present results are in many—but not all—ways in line with a propositional account of EC. The fact that contingency awareness emerged as, by far, the most important moderator of EC supports the core assumption of propositional models that a relation between a CS and a US can influence the liking of the CS only after a conscious proposition about the CS–US relation has been formed. The impact of several other moderators (e.g., age of the participants, subliminal US presentations, instructions to evaluate CSs spontaneously) can be explained if these moderators are assumed to influence the probability that participants form conscious propositions about CS–US relations. The finding that EC is driven primarily by CS–US co-occurrences was not predicted on the basis of propositional models of EC but is also not incompatible with those models. It is possible, in principle, that EC depends on the formation of propositions about the fact that the CS and the US co-occur rather than propositions about the fact that the CS is a reliable predictor of the US. However, such an explanation is clearly post hoc. It thus reveals the current limitations of propositional models of EC. At present, these models are, in essence, restricted to the assumption that EC and other forms of Pavlovian conditioning should depend on the nonautomatic formation and evaluation of propositions about the CS–US relation. Exactly how propositions are formed and evaluated, what the content of the propositions should be, and how the propositions lead to EC (on the different levels of EC assessment) is not made explicit.

Taken together, the findings of the present meta-analysis yield relatively strong support for the notion that EC is substantially influenced by higher order, propositional processes. Nonetheless, the present findings do not rule out that lower order automatic link– formation mechanisms contribute to EC over and above the consciously formed beliefs about CS–US contingencies. A current issue of vigorous debate is whether the concept of automatic link formation bears enough explanatory value: Should it be incorporated as an important learning mechanism next to propositional learning (resulting in a dual-process framework) or is dismissal of the concept of automatic link formation altogether justified (see Mitchell et al., 2009a, 2009b, and associated commentaries)? A full discussion of this thorny issue is clearly beyond the scope of this article. We wish to point out, however, that whether a dual-process account or a single (propositional) account should be preferred depends on metatheoretical issues such as whether priority is given to parsimony or explanatory power. If parsimony is key, the present meta-analytic evidence clearly favors adopting a propositional account while trading off explanatory power with respect to the full range of findings. A dual-process account, by contrast, is less simplistic but seems to be able to account for all of the meta-analytic evidence at hand. Such an account carries some additional problems as well as possibilities: A potential problem is that dual-process models may be more difficult to falsify than single account models, especially when no assumptions are made about how the systems interact (e.g., Mitchell et al., 2009a). An attractive possibility on which we elaborate below is that a dual-systems account can lead to a new, metaconditional approach in EC research. With such an approach, the focus of research is shifted toward identifying the specific conditions under which one process is more likely to be at work than the other rather than on proving whether a particular single process always underlies EC.

Limitations and Future Directions

Like all meta-analyses, the present work is limited by the empirical evidence available and by methodological constraints. First, even though substantial efforts were made to collect all of the available evidence to date, several issues could not be examined in detail because of the small number of studies addressing those issues. For instance, there were probably not enough studies to allow for a solid estimation of the effects of CS subliminal presentation and of a range of special designs. These issues warrant further scrutiny from a meta-analytic perspective as the number of primary studies addressing these issues increases.

Second, meta-analytic estimates can be influenced by the presence of publication bias. Although a general publication bias seems to be unlikely for the present analysis given the results of the Egger test (regression of effect size on sample size) and the graphical inspection of the data reported earlier, there is one puzzling finding that may be indicative of a more subtle form of selection bias: Effect sizes for LD contrasts were not much (and not significantly) larger than LN and DN contrasts even though, on logical grounds, the former should have approximately twice the size of the latter (see Figure 2). One likely explanation in terms of selection bias is that smaller EC effects can be reliably detected by including and reporting LD contrasts instead of LN and DN contrasts. A second possible explanation is that, LN and DN designs may often be subject to a contrast effect whereby the neutral CS acquires, to a certain extent, the opposite valence of the CS paired with the valenced US, thus becoming less "neutral" than expected. Irrespective of whether these two explanations can fully account for the absence of an effect, it is reassuring that an additional analysis involving only LN and DN contrasts as part of our sensitivity analysis produced largely identical results and conclusions in comparison with the main model of our analysis, which included all three effect size contrasts.

Third, we could only indirectly account for possible covariation between moderators. The reasons were methodological in nature and had to do with the aggregation of effect sizes within studies and a considerable loss in power when considering moderators simultaneously. However, unless otherwise noted, the degree of interdependence among coded moderators was reassuringly low. Therefore, strong confounds seem to be unlikely in most cases. Nevertheless, the present meta-analytic results should not be overinterpreted in causal terms. They are best viewed as reference points that help to summarize and structure the available empirical evidence to date.

In a related vein, perhaps the most serious limitation of the current meta-analysis is that we were unable to examine interactions between different moderators. Analyzing such interactions by meta-analysis would become easier if primary research directly investigated such interactions. Until recently, researchers implicitly or explicitly assumed that EC is a unitary phenomenon that is always driven by the same process (e.g., association formation or propositional processes). Given such a view, a potential moderator should have a consistent effect across different instantiations of EC. Although the effect of a moderator on EC might vary from study to study because of random variance, researchers eventually should be able to estimate its "true" moderator effect. Recently, researchers started realizing that different EC effects might be due to different processes (De Houwer et al., 2005; De Houwer, 2007a). If this is actually the case, the effect of certain moderators might depend on the type of process that produces a particular EC effect. De Houwer (2007a), therefore, urged researchers to adopt a metaconditional approach as a next major step for the study of EC. Rather than examining the effect of a single moderator on EC, studies should be directed toward examining whether the effect of a particular moderator depends on other potential moderators. For instance, it might be the case that automatic associative processes can lead to EC if participants are discouraged from consciously detecting CS–US relations and encouraged to evaluate CSs in a spontaneous manner. Under these conditions, EC might be independent of contingency awareness. When participants are encouraged to detect CS–US relations and to justify their evaluations in a rational manner, EC might be due to the formation of propositions. In that case, EC should depend on contingency awareness. Hence, rather than trying to find out whether EC depends on contingency awareness in general, metaconditional research aims to discover when EC depends on contingency awareness. Once a sufficient number of metaconditional studies have been conducted, a new meta-analysis of EC research could again provide important insights into the fundamental question of where our likes and dislikes come from.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Allen, C. T., & Janiszewski, C. A. (1989). Assessing the role of contingency awareness in attitudinal conditioning with implications for advertising research. *Journal of Marketing Research, 26,* 30–43.

*Allen, C. T., & Madden, T. J. (1985). A closer look at classical conditioning. *Journal of Consumer Research, 12,* 301–315.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (Vol. 2, pp. 798–844). Worcester, MA: Clark University Press.

Baeyens, F., & Crombez, G. (1994). Twenty years of evaluative conditioning research: A meta-evaluation. *Unpublished manuscript.*

*Baeyens, F., Crombez, G., Hendrickx, H., & Eelen, P. (1995). Parameters of human evaluative flavor–flavor conditioning. *Learning and Motivation, 26,* 141–160.

*Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact, always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10,* 179–199.

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). *Behaviour Research and Therapy, 33,* 825–831.

*Baeyens, F., Eelen, P., Crombez, G., & De Houwer, J. (2001). On the role of beliefs in observational flavor conditioning. *Current Psychology, 20,* 183–203.

*Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30,* 133–142.

*Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective–evaluative learning. *Cognition & Emotion, 4,* 3–18.

*Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989a). Acquired affective evaluative value: Conservative but not unchangeable. *Behaviour Research and Therapy, 27,* 279–287.

*Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989b). The influence of CS–UCS perceptual similarity/dissimilarity on human evaluative learning and signal learning. *Learning and Motivation, 20,* 322–333.

*Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1990). Flavor–flavor and color–flavor conditioning in humans. *Learning and Motivation, 21,* 434–455.

*Baeyens, F., Hermans, D., & Eelen, P. (1993). The role of CS–US contingency in human evaluative conditioning. *Behaviour Research and Therapy, 31,* 731–737.

*Baeyens, F., Kaes, B., Eelen, P., & Silverans, P. (1996). Observational evaluative conditioning of an embedded stimulus element. *European Journal of Social Psychology, 26,* 15–28.

*Baeyens, F., Vansteenwegen, D., De Houwer, J., & Crombez, G. (1996). Observational conditioning of food valence in humans. *Appetite, 27, 235*–250.

Baeyens, F., Vansteenwegen, D., & Hermans, D. (2009). Associative learning requires associations, not propositions. *Behavioral and Brain Sciences, 32,* 198–199.

*Baeyens, F., Wrzesniewski, A., De Houwer, J., & Eelen, P. (1996). Toilet rooms, body massages, and smells: Two field studies on human evaluative odor conditioning. *Current Psychology, 15,* 77–96.

*Berkowitz, L., & Knurek, D. A. (1969). Label-mediated hostility generalization. *Journal of Personality and Social Psychology, 13,* 200–206.

*Bernstein, I. L. (1978). Learned taste aversions in children receiving chemotherapy. *Science, 200,* 1302–1303.

*Bernstein, I. L., & Webster, M. W. (1980). Learned taste aversions in humans. *Physiology and Behavior, 25,* 363–366.

*Blair, M. E., & Shimp, T. A. (1992). Consequences of an unpleasant experience with music: A second-order negative conditioning perspective. *Journal of Advertising, 21,* 35–43.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Sussex, United Kingdom: Wiley.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11,* 485–494.

*Brunstrom, J. M., & Higgs, S. (2002). Exploring evaluative conditioning using a working memory task. *Learning and Motivation, 33,* 433–455.

*Byrne, D., & Clore, G. L. (1970). A reinforcement model of evaluative responses. *Personality: An International Journal, 1,* 103–128.

*Cacioppo, J. T., Marshall-Goodell, B. S., Tassinary, L. G., & Petty, R. E. (1992). Rudimentary determinants of attitudes: Classical conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology, 28,* 207–233.

*Campoy Menendez, G., Romero Medina, A., & Martinez Sanchez, F. (1997). Evaluative conditioning with subliminal presentations of unconditioned stimuli: An experimental study. *Anales de Psicologia, 13,* 177–187.

*Cicero, S. D., & Tryon, W. W. (1989). Classical conditioning of meaning: II. A replication and triplet associative extension. *Journal of Behavior Therapy and Experimental Psychiatry, 20,* 197–202.

*Cohen, B. H. (1964). Role of awareness in meaning established by classical conditioning. *Journal of Experimental Psychology, 67,* 373–378.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Sage.

*Coppens, E., Vansteenwegen, D., Baeyens, F., Vandenbulcke, M., Van Paesschen, W., & Eelen, P. (2006). Evaluative conditioning is intact after unilateral resection of the anterior temporal lobe in humans. *Neuropsychologia, 44,* 840–843.

*Covin, R. (2008). *Changing self-schema organization using an evaluative conditioning procedure.* Doctoral dissertation, University of Western Ontario.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170–180.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163–170.

*Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology, 89,* 129–142.

Davey, G. C. L. (1994a). Defining the important theoretical questions to ask about evaluative conditioning: A reply to Martin and Levey (1994). *Behaviour Research and Therapy, 32,* 307–310.

Davey, G. C. L. (1994b). Is evaluative conditioning a qualitatively distinct form of classical conditioning? *Behaviour Research and Therapy, 32,* 291–299.

*Dawson, M. E., Rissling, A. J., Schell, A. M., & Wilcox, R. (2007). Under what conditions can human affective conditioning occur without contingency awareness? Test of the evaluative conditioning paradigm. *Emotion, 7,* 755–766.

Dawson, M. E., & Schell, A. M. (1985). Information processing and human autonomic classical conditioning. In P. K. Ackles, J. R. Jennings, & M. Coles (Eds.), *Advances in psychophysiology* (Vol. 1). Greenwich, CT: JAI Press.

*De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation, 37,* 176–187.

De Houwer, J. (2007a). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230–241.

*De Houwer, J. (2007b). *CS-only.* Unpublished manuscript.

*De Houwer, J. (2007c). *CS-only 2.* Unpublished manuscript.

*De Houwer, J. (2007d). *Replica.* Unpublished manuscript.

De Houwer, J. (2009a). *Leerpsychologie* [Learning Psychology]. Gent, Belgium: Academia Press.

De Houwer, J. (2009b). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37,* 1–20.

De Houwer, J. (in press). Evaluative conditioning: A review of procedure knowledge and mental process theories. In T. Schachtman & Steve Reilly (Eds.), *Applications of learning and conditioning.* Oxford, United Kingdom: Oxford University Press.

*De Houwer, J., Baeyens, F., & Eelen, P. (1994). Verbal evaluative conditioning with undetected US presentations. *Behaviour Research and Therapy, 32,* 629–633.

De Houwer, J., Baeyens, F., & Field, A. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19,* 161–174.

*De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture–picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 26,* 237–242.

*De Houwer, J., Beckers, T., & Moors, A. (2006). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43,* 972–978.

*De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with “subliminally” presented stimuli. *Consciousness and Cognition, 6,* 87–107.

*De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition & Emotion, 12,* 145–169.

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

*Díaz, E., & De la Casa, L. G. (2002). Latent inhibition in human affective learning. *Emotion, 2,* 242–250.

*Díaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19,* 245–268.

*Dickinson, A., & Brown, K. J. (2007). Flavor-evaluative conditioning is unaffected by contingency knowledge during training with color-flavor compounds. *Learning & Behavior, 35,* 36–42.

*Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86,* 345–355.

*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. *Learning & Memory, 11,* 549–554.

Dunlap, W. P., Cortina, J. M., Valsow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

*Dwyer, D. M. (2007). *Differences in the effects of cue competition in evaluative and predictive judgments.* Unpublished manuscript.

*Dwyer, D. M., Jarratt, F., & Dick, K. (2007). Evaluative conditioning with foods as CSs and body shapes as USs: No evidence for sex differences, extinction, or overshadowing. *Cognition & Emotion, 21,* 281–299.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634.

*Eifert, G. H., Craill, L., Carey, E., & O'Connor, C. (1988). Affect modification through evaluative conditioning with music. *Behaviour Research and Therapy, 26,* 321–330.

Estes, W. K. (1997). On the communication of information by displays of standard errors and confidence intervals. *Psychonomic Bulletin and Review, 4,* 330–341.

Everhart, D. E., & Demaree, H. A. (2003). Low alpha power (7.5–9.5 Hz) changes during positive and negative affective learning. *Cognitive, Affective & Behavioral Neuroscience, 3,* 39–45.

*Field, A. P. (1997). *Re-evaluating evaluative conditioning.* Doctoral dissertation, University of Sussex, Brighton, United Kingdom.

*Field, A. P. (1998). *Visual EC drug study.* Unpublished manuscript.

Field, A. P. (2005). Learning to like (or dislike): Associative learning of preferences. In A. J. Wills (Ed.), *New directions in human associative learning* (pp. 221–252). Mahwah, NJ: Erlbaum.

*Field, A. P., & Brook, R. (1999). *Evaluative conditioning: Reassessing the role of contingency awareness using a counterbalanced paradigm.* Unpublished manuscript.

*Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology: Animal Behavior Processes, 25,* 211–224.

*Field, A. P., & Garcher, H. (2002). *An investigation into evaluative conditioning using fear relevant stimuli.* Unpublished manuscript.

*Field, A. P., & Houseman, L. (2007). *Is negative affect a boundary condition of evaluative conditioning?* Unpublished manuscript.

*Field, A. P., & Lucas, K. (1999). *The role of celebrity and personal information on advertising.* Unpublished manuscript.

*Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion, 19,* 217–243.

*Field, A. P., & Reed, K. (2006). *Does evaluative conditioning exist when using subliminally presented stimuli?* Unpublished manuscript.

*Field, A. P., & Sutcliffe, J. (2005). *Contingency awareness in evaluative conditioning.* Unpublished manuscript.

*Field, A. P., Tree, J., & Davey, G. C. L. (1998). *Evaluative conditioning using disgust facial expressions as unconditioned stimuli: Evidence for selective conditioning effects.* Unpublished manuscript.

*Field, A. P., & Ugland, C. C. O. (2005). *Does a conceptual connection between the CS and US facilitate visual evaluative conditioning?* Unpublished manuscript.

Fox, E. (2009). *Emotion science.* Basingstoke, United Kingdom: Macmillan.

*Fulcher, E. P., & Hammerl, M. (2001). When all is revealed: A dissociation between evaluative learning and contingency awareness. *Consciousness and Cognition, 10,* 524–549.

*Fulcher, E. P., Mathews, A., Mackintosh, B., & Law, S. (2001). Evaluative learning and the allocation of attention to emotional stimuli. *Cognitive Therapy and Research, 25,* 261–280.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731.

*Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology, 44,* 1355–1361.

*Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research, 35,* 178–188.

*Glaser, J. C. (1999). *The relation between stereotyping and prejudice: Measures of newly formed automatic associations.* Doctoral dissertation, Department of Psychology, Yale University.

*Gorn, G. J. (1982). The effects of music in advertising on choice behavior: A classical conditioning approach. *Journal of Marketing, 46,* 94–101.

Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (Vol. 383–409). New York: Russell Sage Foundation.

*Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90,* 1–20.

*Groenland, E. A. G., & Schoormans, J. P. L. (1994). Comparing mood-induction and affective conditioning as mechanisms influencing product evaluation and product choice. *Psychology & Marketing, 11,* 183–197.

*Grossman, R. P., & Till, B. D. (1998). The persistence of classically conditioned brand attitudes. *Journal of Advertising, 27,* 23–31.

*Hamm, A. O., Greenwald, M. K., Bradley, M. M., & Lang, P. J. (1993). Emotional learning, hedonic change, and the startle probe. *Journal of Abnormal Psychology, 102,* 453–465.

*Hamm, A. O., & Vaitl, D. (1996). Affective learning: Awareness and aversion. *Psychophysiology, 33,* 698–710.

*Hammerl, M., Bloch, M., & Silverthorne, C. P. (1997). Effects of US-alone presentations on human evaluative conditioning. *Learning and Motivation, 28,* 491–509.

*Hammerl, M., & Fulcher, E. P. (2005). Reactance in affective–evaluative learning: Outside of conscious control? *Cognition & Emotion, 19,* 197–216.

*Hammerl, M., & Grabitz, H.-J. (1996). Human evaluative conditioning without experiencing a valued event. *Learning and Motivation, 27,* 278–293.

*Hammerl, M., & Grabitz, H.-J. (2000). Affective–evaluative learning in humans: A form of associative learning or only an artifact? *Learning and Motivation, 31,* 345–363.

*Hardwick, S., & Lipp, O. (2000). Modulation of affective learning: An occasion for evaluative conditioning? *Learning and Motivation, 31,* 251–271.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Hedges, L. V. (1994). Fixed-effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (Vol. 285–299). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Herman, C. P., & Polivy, J. (2004). The self-regulation of eating. In R. F. Baumeister & K. D. Vohs (Eds.), *The handbook of self-regulation: Research, theory, and applications* (pp. 492–508). New York: Guilford Press.

*Hermans, D., Baeyens, F., Lamote, S., Spruyt, A., & Eelen, P. (2005). Affective priming as an indirect measure of food preferences acquired through odor conditioning. *Experimental Psychology, 52,* 180–186.

*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 S. P. Shohov (Ed.), *Advances in psychology research* (Vol. 12, pp. 17–40). Hauppauge, NY: Nova Science.

*Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2004). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43,* 533–551.

*Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion, 17,* 83–99.

*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. *Behaviour Research and Therapy, 40,* 217–234.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327,* 557–560.

*Hill, T., Lewicki, P., Czyzewska, M., & Boss, A. (1989). Self-perpetuating development of encoding biases in person perception. *Journal of Personality and Social Psychology, 57,* 373–387.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385.

Hofmann, W., Gschwendner, T., Nosek, B., & Schmitt, M. (2005). What moderates implicit–explicit consistency? *European Review of Social Psychology, 16,* 335–390.

*Huijding, J., Bos, A. E. R., & Muris, P. (2007). *Enhancing implicit self-esteem in children.* Unpublished manuscript.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

*Huntjens, R. J. C., Peters, M. L., Postma, A., Woertman, L., Effting, M., & van der Hart, O. (2005). Transfer of newly acquired stimulus valence between identities in dissociative identity disorder (DID). *Behaviour Research and Therapy, 43,* 243–255.

*Hvasta, L., & Zanuttini, L. (1989). Odour memory and odour hedonics in children. *Perception, 18,* 391–396.

Jaanus, H., Defares, P. B., & Zwaan, E. J. (1990). Verbal classical conditioning of evaluative responses. *Advances in Behaviour Research & Therapy, 12,* 123–151.

*Jarvis, W. B. G. (1998). *Do attitudes really change?* Doctoral dissertation, Ohio State University.

*Johnsrude, I. S., Owen, A. M., Zhao, W. V., & White, N. M. (1999). Conditioned preference in humans: A novel experimental approach. *Learning and Motivation, 30,* 250–264.

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology, 96,* 933–948.

*Kellaris, J. J., & Cox, A. D. (1989). The effects of background music in advertising: A reassessment. *Journal of Consumer Research, 16,* 113–118.

*Kim, H.-J. (2006). *Attitude formation in compulsive hoarding.* Doctoral dissertation, Boston University.

*Kim, J., Lim, J.-S., & Bhargava, M. (1998). The role of affect in attitude formation: A classical conditioning approach. *Journal of the Academy of Marketing Science, 26,* 143–152.

*Krosnick, J. A., Betz, A. L., Jussim, L. J., & Lynn, A. R. (1992). Subliminal conditioning of attitudes. *Personality and Social Psychology Bulletin, 18,* 152–162.

Lang, P. M. (1995). The emotion probe: Studies on motivation and attention. *American Psychologist, 50,* 372–385.

*Lascelles, K. R. R., & Davey, G. C. L. (2006). Successful differential evaluative conditioning using simultaneous and trace conditioning procedures in the picture–picture paradigm. *The Quarterly Journal of Experimental Psychology, 59,* 482–492.

*Lascelles, K. R. R., Field, A. P., & Davey, G. C. L. (2003). Using foods as CSs and body shapes as UCSs: A putative role for associative learning in the development of eating disorders. *Behavior Therapy, 34,* 213–235.

Levey, A. B., & Martin, I. (1975). Classical conditioning of human 'evaluative' responses. *Behaviour Research and Therapy, 4,* 205–207.

Levey, A. B., & Martin, I. (1990). Evaluative conditioning: Overview and further options. *Cognition & Emotion, 4,* 31–37.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

*Lipp, O. V., Cox, D., & Siddle, D. A. T. (2001). Blink startle modulation during anticipation of pleasant and unpleasant stimuli. *Journal of Psychophysiology, 15,* 155–162.

*Lipp, O. V., Neumann, D. L., & Mason, V. (2001). Stimulus competition in affective and relational learning. *Learning and Motivation, 32,* 306–331.

*Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34,* 219–239.

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition & Emotion, 19,* 269–282.

*Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 431–440.

*Lipp, O. V., Siddle, D. A. T., & Dall, P. J. (2003). The effects of unconditional stimulus valence and conditioning paradigm on verbal, skeleto-motor and autonomic indices of human Pavlovian conditioning. *Learning and Motivation, 34,* 32–51.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*Livingston, R. W., & Drwecki, B. B. (2007). Why are some individuals not racially biased? Susceptibility to affective conditioning predicts nonprejudice toward Blacks. *Psychological Science, 18,* 816–823.

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

*Mallan, K. M., & Lipp, O. V. (2007). Does emotion modulate the blink reflex in human conditioning? Startle potentiation during pleasant and unpleasant cues in the picture–picture paradigm. *Psychophysiology, 44,* 737–748.

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy, 1,* 57–102.

Martin, I., & Levey, A. B. (1994). The evaluative response: Primitive but necessary. *Behaviour Research and Therapy, 32,* 301–305.

*Michael, T., & Vriends, N. I. (2007). *I am scared of it, but I don't know why: Conditioning without contingency awareness.* Unpublished manuscript.

*Michael, T., Vriends, N., Blechert, J., Meyer, A. H., Zetsche, U., Margraf, J., & Wilhelm, F. H. (2007). *Strong associative learning between neutral and trauma-related stimuli in posttraumatic stress disorder in the absence of contingency awareness.* Unpublished manuscript.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009a). Authors' response: Link-based learning theory creates more problems than it solves. *Behavioral and Brain Sciences, 32,* 230–238.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009b). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,* 183–198.

*Mullin, T. A. (2004). *Using pills as CSs and body shapes as UCSs: Does associative learning play a role in the development of eating disorders?* Unpublished manuscript.

Nelson, J. P., & Kennedy, P. E. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental and Resource Economics, 42,* 345–377.

*Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology, 26,* 505–527.

O'Donnell, J. M., & Brown, M. J. K. (1973). The classical conditioning of attitudes. A comparative study of ages 8 to 18. *Journal of Personality and Social Psychology, 26,* 379–385.

*Olatunji, B. O. (2006). Evaluative learning and emotional responding to fearful and disgusting stimuli in spider phobia. *Journal of Anxiety Disorders, 20,* 858–876.

*Olatunji, B. O., Lohr, J. M., Sawchuk, C. N., & Westendorf, D. H. (2005). Using facial expressions as CSs and fearsome and disgusting pictures as UCSs: Affective responding and evaluative learning of fear and disgust in blood-injection-injury phobia. *Journal of Anxiety Disorders, 19,* 539–555.

*Olson, M. A. (2004). *Reducing automatically-activated racial prejudice through implicit evaluative conditioning.* Doctoral dissertation, Indiana University.

*Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12,* 413–417.

*Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition, 20,* 89–104.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354–379.

*Page, M. M. (1974). Demand characteristics and the classical conditioning of attitudes experiment. *Journal of Personality and Social Psychology, 30,* 468–476.

*Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: Social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology, 59,* 475–486.

*Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 130–144.

*Pleyers, G., Corneille, O., Yzerbyt, V., & Luminent, O. (2008). *Evaluative conditioning may incur attentional costs.* Unpublished manuscript.

Poulton, R., & Menzies, R. G. (2002). Fears born and bred: Toward a more inclusive theory of fear acquisition. *Behaviour Research and Therapy, 40,* 197–208.

*Prestwich, A., Perugini, M., Hurling, R., & Richetin, J. (2010). Using the self to change implicit attitudes. *European Journal of Social Psychology, 40,* 61–71.

*Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation, 32,* 84–99.

Razran, G. (1954). The conditioned evocation of attitudes (cognitive conditioning?). *Journal of Experimental Psychology, 48,* 278–282.

*Redker, C. M., & Gibson, B. (2009). Music as an unconditioned stimulus: Positive and negative effects of country music on implicit attitudes, explicit attitudes, and brand choice. *Journal of Applied Social Psychology, 31,* 2689–2705.

Rescorla, R. A. (1966). Predictability and number of pairings in Pavlovian fear conditioning. *Psychonomic Science, 4,* 383–384.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist, 43,* 151–160.

Rosenthal, D. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin, 92,* 500–504.

Rouder, J. N., & Morey, R. D. (2005). Relational and arelational confidence intervals. *Psychological Science, 16,* 77–79.

Rozin, P. (1982). Human food selection: The interaction of biology, culture and individual experience. In L. M. Barker (Ed.), *The psychobiology of human food selection* (pp. 225–254). Westport, CT: Avi.

Rozin, P., & Millman, L. (1987). Family environment, not heredity, accounts for family resemblances in food preferences and attitudes: A twin study. *Appetite, 8,* 125–134.

*Rozin, P., Wrzesniewski, A., & Byrnes, D. (1998). The elusiveness of evaluative conditioning. *Learning and Motivation, 29,* 397–415.

Russell, J. A. (2003). Core affect and the psychological construct of emotion. *Psychological Review, 110,* 145–172.

*Sachs, D. H., & Byrne, D. (1970). Differential conditioning of evaluative responses to neutral stimuli through association with attitude statements. *Journal of Experimental Research in Personality, 4,* 181–185.

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62,* 97–128.

*Schneider, F., Weiss, U., Kessler, C., Muller-Gartner, H.-W., Posse, S., Salloum, J. B., et al. (1999). Subcortical correlates of differential classical conditioning of aversive emotional reactions in social phobia. *Biological Psychiatry, 45,* 863–871.

Sedlmeier, P., & Renkewitz, F. (2007). *Forschungsmethoden und Statistik in der Psychologie* [Research methods and statistics in psychology]. München, Germany: Pearson Studium.

*Shimp, T. A., Stuart, E. W., & Engle, R. W. (1991). A program of classical conditioning experiments testing variations in the conditioned stimulus and context. *Journal of Consumer Research, 18,* 1–12.

*Spruyt, A., Hermans, D., De Houwer, J., & Eelen, P. (2004). Automatic non-associative semantic priming: Episodic affective priming of naming responses. *Acta Psychologica, 116,* 39–54.

*Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *The Journal of Abnormal and Social Psychology, 57,* 37–40.

*Staats, A. W., Staats, C. K., & Biggs, D. A. (1958). Meaning of verbal stimuli changed by conditioning. *American Journal of Psychology, 71,* 429–431.

*Staats, A. W., Staats, C. K., Finley, J. R., & Minke, K. A. (1963). Meaning established by classical conditioning controlling associates to the UCS. *Journal of General Psychology, 69,* 247–252.

*Staats, A. W., Staats, C. K., Heard, W. G., & Nims, L. P. (1959). Meaning established by classical conditioning. *Journal of Experimental Psychology, 57,* 64.

*Staats, C. K., & Staats, A. W. (1957). Meaning established by classical conditioning. *Journal of Experimental Psychology, 54,* 74–80.

*Stahl, C., & Degner, J. (2007). Assessing automatic activation of valence: A multinomial model of EAST performance. *Experimental Psychology, 54,* 99–112.
*Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple USs: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 286–291.
*Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of US valence and US Identity in valence acquisition through evaluative conditioning. *Journal of Personality and Social Psychology, 97,* 404–420.
*Stevenson, R. J., & Prescott, J. (1995). The acquisition of taste properties by odors. *Learning and Motivation, 26,* 433–455.
*Strick, M., Holland, R. W., & van Knippenberg, A. (2008). Seductive eyes: Attractiveness and direct gaze increase desire for associated objects. *Cognition, 106,* 1487–1496.
*Strick, M., van Baaren, R. B., Holland, R. W., & van Knippenberg, A. (2009). Humor in advertisements enhances product liking by mere association. *Journal of Experimental Psychology: Applied, 15,* 35–45.
*Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical conditioning of consumer attitudes: Four experiments in an advertising context. *Journal of Consumer Research, 14,* 334–349.
*Theis, F., Boakes, R. A., Karazinov, D., & Walther, E. (2006). *Would you like him if he took a shower? Changing evaluations of faces paired with odors.* Unpublished manuscript.
*Todrank, J., Byrnes, D., Wrzesniewski, A., & Rozin, P. (1995). Odors can change preferences for people in photographs: A cross-modal evaluative conditioning study with olfactory USs and visual CSs. *Learning and Motivation, 26,* 116–140.
*Tryon, W. W., & Briones, R. G. (1985). Higher-order semantic counterconditioning of Filipino women's evaluations of heterosexual behaviors. *Journal of Behavior Therapy and Experimental Psychiatry, 16,* 125–131.
*Tryon, W. W., & Cicero, S. D. (1989). Classical conditioning of meaning: I. A replication and higher-order extension. *Journal of Behavior Therapy and Experimental Psychiatry, 20,* 137–142.
*Van Reekum, C. M., van den Berg, H., & Frijda, N. H. (1999). Cross-modal preference acquisition: Evaluative conditioning of pictures by affective olfactory and auditory cues. *Cognition & Emotion, 13,* 831–836.
*Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (1998). Extinction in fear conditioning: Effects on startle modulation and evaluative self-reports. *Psychophysiology, 35,* 729–736.
*Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 71–79.
*Verhulst, F., Hermans, D., Baeyens, F., Spruyt, A., & Eelen, P. (2006). Determinants and predictive validity of direct and indirect measures of recently acquired food attitudes. *Appetite, 46,* 137–143.
*Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82,* 919–934.
*Walther, E., & Grigoriadis, S. (2004). Why sad people like shoes better: The influence of mood on the evaluative conditioning of consumer attitudes. *Psychology & Marketing, 21,* 755–773.
*Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 454–459.
*Wardle, S. G., Mitchell, C. J., & Lovibond, P. F. (2007). Flavor evaluative conditioning and contingency awareness. *Learning & Behavior, 35,* 233–241.
*Whitfield, M., & Jordan, H. (2007). *Mutual influence of implicit and explicit attitudes.* Unpublished manuscript.
*Yeomans, M. R., Mobini, S., Elliman, T. D., Walker, H. C., & Stevenson, R. J. (2006). Hedonic and sensory characteristics of odors conditioned by pairing with tastants in humans. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 215–228.

Received May 12, 2009
Revision received October 30, 2009
Accepted December 4, 2009 ■