A systematic review and meta-analysis of the evidence for unaware fear conditioning

Mertens, Gaetan; Engelhard, Iris M.

Review article

A systematic review and meta-analysis of the evidence for unaware fear conditioning

Gaëtan Mertens*, Iris M. Engelhard

Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands

ABSTRACT

Whether fear conditioning can take place without contingency awareness is a topic of continuing debate and conflicting findings have been reported in the literature. This systematic review provides a critical assessment of the available evidence. Specifically, a search was conducted to identify articles reporting fear conditioning studies in which the contingency between conditioned stimuli (CS) and the unconditioned stimulus (US) was masked, and in which CS-US contingency awareness was assessed. A systematic assessment of the methodological quality of the included studies ($k = 41$) indicated that most studies suffered from methodological limitations (i.e., poor masking procedures, poor awareness measures, researcher degrees of freedom, and trial-order effects), and that higher quality predicted lower odds of studies concluding in favor of contingency unaware fear conditioning. Furthermore, meta-analytic moderation analyses indicated no evidence for a specific set of conditions under which contingency unaware fear conditioning can be observed. Finally, funnel plot asymmetry and *p*-curve analysis indicated evidence for publication bias. We conclude that there is no convincing evidence for contingency unaware fear conditioning.

1. Introduction

Classical conditioning is one of the oldest and most established procedures within psychology. In this procedure, two different stimuli, the conditioned stimulus (CS) and the unconditioned stimulus (US), are paired, resulting in conditioned responses (CRs) to the CS (Pavlov, 1928). A variant, the fear conditioning procedure, involves the pairing of usually an initial neutral CS with an aversive US (e.g., an electric shock) (Lonsdorf et al., 2017). Established CRs involve a range of different behaviors related to fear, such as elevated skin conductance responses, potentiated startle responses, subjective distress, and avoidance behavior. The fear conditioning procedure is regarded as an important model for the etiology of anxiety and fear-related disorders (Mineka and Zinbarg, 2006; Rachman, 1991; Vervliet et al., 2013).

The processes underlying classical conditioning, including fear conditioning, are often seen as simple and automatic. For example, in the Encyclopedia of Human Behavior, classical conditioning is introduced as a "natural phenomenon of reflexive learning" (Ploog, 2012, p. 484). However, these assumptions have been repeatedly challenged by prominent authors in the conditioning field, who have highlighted that the reflexive system is an unlikely explanation of several findings with classical learning procedures, such as the role of contingency rather than contiguity, preparatory rather than reflexive conditioned

responses, the involvement of inferential reasoning, and effects of verbal instructions (Dawson and Furedy, 1976; Grings, 1973; Lovibond and Shanks, 2002; Mertens et al., 2018; Rescorla, 1988). Yet some of the most influential contemporary models about human learning and memory maintain that classical conditioning can take place entirely through automatic learning processes (Amodio, 2018; LeDoux, 2014; Squire, 2004). One of the strongest and most persistent assumptions about classical (including fear) conditioning is that it can take place in the absence of awareness of the CS-US contingency (which is one of several features of automaticity; see Bargh, 1994; McNally, 1995; Moors and De Houwer, 2006).

The idea that fear conditioning can take place without contingency awareness was introduced early in the literature (Diven, 1937; Haggard, 1943; Lacey and Smith, 1954). However, other studies quickly followed, challenging these findings and demonstrating that fear conditioning only occurs when participants are able to verbalize the CS-US contingency (indicating CS-US contingency awareness) (Chatterjee and Eriksen, 1960, 1962, Fuhrer and Baer, 1965, 1969). Still, the question remains largely unresolved: many empirical reports and reviews have argued that fear conditioning does not occur in the absence of CS-US awareness (Brewer, 1974; Dawson and Furedy, 1976; Lovibond and Shanks, 2002; Mitchell et al., 2009), while a comparable series of empirical reports and theoretical papers have argued that it

does occur without awareness (Öhman and Mineka, 2001; Schultz and Helmstetter, 2010; Sevenster et al., 2014; Weike et al., 2007). This status of the debate may be attributed to methodological and (meta-)theoretical issues that complicate the evaluation of the evidence. We will first provide a brief overview of these issues, and they will serve as a benchmark for our systematic review below.

1.1. Methodological problems regarding contingency unaware fear conditioning

1.1.1. Masking procedures

Studies investigating unaware fear conditioning usually employ some type of masking procedure to prevent the development of contingency awareness. Most commonly, the CS-US contingency is masked by interfering with the perception of the CS (e.g., backward masking, continuous flash suppression). If perception of the CS is successfully suppressed, it can be argued that the CS-US contingency is also masked (although perception of CS may not be completely suppressed on all feature levels and pseudo-random trial-order can be used to develop partial knowledge of the CS-US contingency; see Sections 1.1.1.3 and 1.1.4). Alternatively, other procedures (e.g., a distractor task, instruction manipulations) can be used to manipulate participants' attention for the contingencies and mask the CS-US contingency, without interfering with the perception of the CS. However, there are potential limitations to these masking procedures.

1.1.1.1. Suboptimal parameters. The effectiveness of a masking procedure depends on whether the parameters of the procedure are well adjusted. For instance, in backward masking, a briefly presented stimulus (often between $10-50$ ms) is followed by a mask (e.g., a grey pattern) to prevent development of awareness of the masked stimulus. However, when the stimulus presentation or the interval between the stimulus and mask are too long, most participants will be able to perceive the stimulus (Vermeiren and Cleeremans, 2012). Similarly, a distractor task (e.g., a n-back task in which participants are asked to report information in a trial that was presented n-trials earlier) may be ineffective if the task is too easy or too hard. Hence, masking procedures require adequate parameters to ensure that they are effective, otherwise a substantial portion of the sample will develop awareness.

1.1.1.2. Individual differences. Individual differences between participants (e.g., regarding eye-sight, working memory capacity, motivation, and attention) may result in less effective masking procedures. For instance, stimuli presented for 33 ms in a masking task may be imperceptible for some participants but clearly visibly for others (Pessoa et al., 2006). Likewise, 2-back tasks (which require participants to report a stimulus presented 2 trials ago) may not be challenging for participants with high working memory capacity. Neglecting such inter-individual differences can result in contingency awareness in part of the sample despite using a masking task. Similarly, *intra*-individual differences, such as variations in motivation and attention (which can, of course, also vary between participants), can render masking procedures less effective. For example, in continuous flash suppression, participants are presented a static stimulus to one eye and a series of rapidly changing stimuli to the other eye, resulting in the suppression of the static stimulus from awareness. However, if participants do not maintain eye fixation, the suppression effect can break (Faivre et al., 2014). Hence, inter-individual (i.e., between participants) and intra-individual (i.e., within participants) differences should be taken into account to ensure that masking procedures are effective. To ensure unawareness, task parameters may be adjusted individually (i.e., for each participant) and dynamically (i.e., throughout the experiment).

1.1.1.3. Presence of other stimulus dimensions. Another potential problem is that participants sometimes use another visual feature (e.g., object size) to broadly categorize masked stimuli. As such, masking procedures may be successful to prevent awareness of stimuli on one feature level, but not the broader classification of stimuli along other perceptual dimensions (Gayet et al., 2019; Gelbard-Sagiv et al., 2016). This may result in a misclassification of participants as unaware, even though they can discriminate stimuli.

1.1.1.4. Solutions. Masking procedures should be used carefully and effectively to ensure that most participants in the sample do not develop awareness. Particularly, to ensure that participants are and remain unaware of stimuli, researchers should use (a) a masking technique; (b) set appropriate parameters for the masking technique (most optimally individually and dynamically adjusted); and (c) ensure that participants cannot use other perceptual dimensions to discriminate between stimuli.

1.1.2. Awareness measures

Given the limitations of masking procedures, manipulation checks (i.e., measures of contingency awareness) are required to select participants who are unaware of the contingencies. Such measures include post-experimental questionnaires, visual discrimination tasks, and trial-by-trial US expectancy ratings. However, these awareness measures also have a number of the previously described limitations (Lovibond and Shanks, 2002; Shanks and St. John, 1994).

1.1.2.1. Insensitive awareness measures. Awareness measures should be sufficiently sensitive to detect contingency awareness, but this can be a problem, for instance, with post-experimental questionnaires. Due to the passage of time, participants may forget relevant information (e.g., the hair color of a face presented earlier), resulting in underperformance on the awareness measure (for such a demonstration see Dawson and Reardon, 1973). Besides avoiding long time-lags, awareness measures should be sufficiently powered. In a review of awareness measures used in implicit learning research, Vadillo et al. (2016) found that behavioral tasks (such as visual discrimination tasks) usually relied on a limited number of trials, resulting in a loss of power and erroneous conclusions of absence of awareness. Hence, it is important to ensure that awareness measures are not affected by factors that may reduce their sensitivity, such as time delays and underpowered tests.

1.1.2.2. Irrelevant awareness measures. Awareness measures should measure knowledge that may indicate (partial) contingency awareness. For instance, as described above, participants may use a different perceptual dimension to discriminate between stimuli than the one the researchers had in mind. If researchers do not assess knowledge that the participants actually used (e.g., "I saw a long, slim object") by asking a question that is too specific or irrelevant (e.g., "Did you see a snake?"), they may erroneously conclude that participants were unaware of the relevant information. This point also relates to controlling for trial-order effect (see Section 1.1.4 below).

1.1.2.3. An appropriate awareness criterium needs to be set. When criteria for awareness are too strict, a high number of participants who are (partially) contingency aware may be classified as 'contingency unaware'. For example, Schultz and Helmstetter (2010) classified participants as contingency unaware when US expectancy ratings (on a 100-point scales) for CS- (i.e., the CS not paired with the US) were rated more than 25 and/or US expectancy ratings for CS + (i.e., the CS paired with the US) were rated as less than 75. This resulted in 8 out of 10 participants in the masking condition being classified as unaware. However, using a more lenient criterion of rating US expectancy for CS- as more than 50 and/or US expectancy for CS + as more less 50, only 4 out of 10 participants were classified as contingency unaware. Hence, an appropriate contingency awareness criterium should be set to

account for possible partial awareness of the contingencies.

1.1.2.4. Regression-to-the-mean. Shanks (2017) reframed these problems relating to awareness measures as a regression-to-the-mean problem. That is, selecting contingency unaware participants on the basis of an awareness measure is a type of extreme group selection (particularly when the masking procedure is not effective, resulting in fewer participants demonstrating the desired awareness level). As such, regression-to-the-mean dictates that participants demonstrating extreme performance on one measure (e.g., no contingency awareness in the awareness test) will tend to show less extreme performance on another measures (i.e., reliable discrimination of the outcome of interest in a fear conditioning procedure). This problem is always present (unless awareness measures and outcome measures are perfectly correlated) and it is particularly pronounced when the masking procedure is not fully effective and when the awareness measure and the outcome of interest are weakly correlated. It can be further noted that awareness measures are typically quite unreliable (Vadillo et al., 2019), which introduces an upper bound for any correlation with these measures (Novick, 1966). Hence, imperfect correlations and regression to the mean are to be expected in all studies addressing unaware processes.

1.1.2.5. Solutions. According to Shanks (2017, one solution to the problem of poor awareness measures and regression-to-the-mean is to use multiple awareness measures. The second awareness measure can be used to independently eliminate bias in the first awareness measure. Furthermore, a second measure can confirm the sensitivity and validity of the first measure. Hence, taken together, measures of awareness should be sensitive and relevant, an appropriate criterium needs to be set, and regression-to-the-mean should be accounted for. A practical solution for these problems is to use two independent measures of awareness.

1.1.3. Researcher degrees of freedom and HARKing

Another challenge for research on contingency unaware fear conditioning is flexibility in the selection, pre-processing, and analysis of the data. Such 'garden forking paths' (Gelman and Loken, 2013) or 'researcher degrees of freedom' (Simmons et al., 2011) increase the risk of finding false-positive results. For instance, fear conditioning studies commonly measure at least one psychophysiological response (e.g., skin conductance response, fear potentiated startle, or heart rate) and one or several subjective measures (e.g., ratings of US expectancy or CS evaluations). Furthermore, skin conductance, the most commonly used outcome measure in fear conditioning, can be scored in different ways (e.g., relative increases using a baseline or trough-to-peak scoring of responses) and in different time intervals (e.g., across the duration of CS presentation or scoring responses in the first and second time interval of CS presentation) (Pineles et al., 2009). Finally, participants are often excluded in fear conditioning research for insufficient data quality. There is a wide range in criteria for excluding participants and these are often not implemented consistently (Lonsdorf et al., 2019). Such flexibility in the analysis of data, without appropriate error control, can inflate the risk of false-positive findings (Gelman and Loken, 2013; Simmons et al., 2011).

The problem of flexibility of data processing can be further exacerbated by cherry picking results across multiple conditions or moderators and flexibility in the interpretation of those obtained results, a practice known as Hypothesizing After the Results are Known (or HARKing) (Kerr, 1998). For instance, when a hypothesis (such as whether or not fear conditioning can occur without contingency awareness) is tested under multiple conditions and with multiple outcome measures, without clear a priori specification of predicted moderators, a researcher could discover an effect in any of the conditions or for any of the measures and selectively present this finding as predicted beforehand. Such practices, in combination with publication bias

favoring positive results (see Section 1.2.2), can inflate the number of spurious findings in the literature (De Groot, 2014; Forstmeier et al., 2017; Kerr, 1998).

1.1.3.1. Solutions. There are ways to mitigate the problems of researcher degrees of freedom and HARKing, such as preregistration of data-analyses plans and hypotheses (Krypotos et al., 2019; van' t Veer and Giner-Sorolla, 2016), but these practices have not yet been routinely been used in fear conditioning research (Lonsdorf et al., 2019). Without pre-specified data-analysis steps and hypotheses, it is difficult to scrutinize whether or not results have been influenced by flexible interpretation and researcher degrees of freedom. Because spurious findings are, per definition, unreliable, a simple (though potentially conservative[1]) decision rule is that inconsistent results across measures and/or conditions might suggest that the results have been potentially influenced by researcher degrees of freedom and/or HARKing.

1.1.4. Trial-order effects

One final potential problem is trial-order effects, which could partially account for the evidence found in favor of unaware fear conditioning. That is, often pseudo-random trial orders are used in fear conditioning research to prevent multiple successive CS-US trials (usually maximally two).[2] This practice introduces a predictable trial order. That is, after an unreinforced trial, the probability of a subsequent reinforced trial is higher (and vice versa; how much precisely depends on the reinforcement rate and the constraints of the pseudo-random trial order). When participants pick up this contingency, they can anticipate the US without necessarily needing to identify the CS. Indeed, several studies (Sevenster et al., 2014; Singh, Dawson et al., 2013; Wiens et al., 2003) have indicated that trial-order effects can account for apparent instances of contingency unaware fear conditioning.

1.1.4.1. Solutions. When a pseudo-random trial order is used, the effects of trial-order on conditioned responses can be investigated by comparing non-alternating trials (i.e., CS + following CS + and CS- following CS-) to alternating trials (i.e., CS + following CS- and vice versa). For alternating trials, trial-order effects will result in reliable discrimination (i.e., expectancy of the US for the CS + compared to the CS-), due to the previous trial, even when participants are completely unaware of the CS-US contingency. To test for contingency unaware fear conditioning, statistical analyses can focus on non-alternating trials. For these trials, participants cannot correctly anticipate (non-)reinforcement on the next trial, and thus evidence for contingency unaware fear conditioning on non-alternating trials cannot be due to trial-order effects.

1.2. Other potential problems with the evidence for contingency unaware fear conditioning

1.2.1. Disagreement about moderators of unaware fear conditioning

Theoretical papers have outlined a number of conditions under which unaware fear conditioning is thought to be reliably observed. Particularly, some researchers have argued that unaware fear

---

[1] Whether or not this decision rule is indeed conservative depends on to what extent the described effects in the literature are reliable. Some authors have argued that most reported research findings reflect false-positive results due to researcher degrees of freedom and HARKing (Ioannidis, 2005; Simmons et al., 2011) and recent replication initiatives in psychology indicate that replicability of research findings is problematic (Klein et al., 2014; Open Science Collaboration, 2015). Therefore, our decision rule may not be that conservative.

[2] Pseudo-random trial orders are used in fear conditioning research to prevent long streaks of reinforced or unreinforced trials, which tend to produce the Gambler's fallacy phenomenon (McAndrew et al., 2012).

conditioning can only be observed with certain outcome variables (i.e., fear potentiated startle, amygdala activity; Hamm and Weike, 2005), with certain procedural parameters (i.e., delay conditioning; Weike et al., 2007), and for certain classes of stimuli (i.e., evolutionary fear-relevant stimuli such as angry male faces and pictures of snakes and spiders; Soares and Öhman, 1993). As of yet, these claims are based on a small number of studies and the results are not necessarily consistent (see Lipp, Kempnich et al., 2014; Lovibond et al., 2011; McNally, 1987). Hence, the evidential basis for these proposed moderators is limited and requires further evaluation.

1.2.2. Publication bias

Publication bias is a well-known problem that can result in the misrepresentation of the robustness and reliability of effects reported in scientific literature (Bakker et al., 2012; Rosenthal, 1979). There are several reasons why the literature on contingency unaware fear conditioning might be affected by publication bias. First, the question of whether fear conditioning can occur without awareness relates to influential ideas about the organization of emotions (i.e., elicitation of emotions requires minimal processing; Damasio, 1994; James, 1884; Zajonc, 1980), memory (i.e., distinction between implicit and explicit memory systems; Squire, 2004), and the etiology of psychopathology (i.e., unaware associations drive maladaptive fear responses; Mineka and Öhman, 2002). These ideas are not uncontested (e.g., Hofmann, 2008; Lazarus, 1982; Moors et al., 2017; Shanks and Berry, 2012), but they remain influential. Articles reporting demonstrations of unaware fear conditioning may have been published more easily because they are in accordance with the predictions of these influential theories (Lee et al., 2013). A second reason why the literature may be affected by publication bias is that demonstrations of an effect are more easily published than negative findings (Coursol and Wagner, 1986; Levine et al., 2009). Third, samples sizes in fear conditioning research are often small, which leads to imprecise and inflated effect size estimates (Button et al., 2013). Finally, a fourth reason that may have contributed to publication bias is that demonstrations of unconscious learning may be perceived as exciting, novel, and unusual results, which are more likely to get published (Nosek et al., 2012; Yong, 2012). Hence, the literature on contingency unaware fear conditioning may be substantially biased, which may have resulted in a misrepresentation of the robustness of this phenomenon.

1.3. Goals of the present systematic review and meta-analysis

This systematic review and meta-analysis will provide an overview of the available evidence from studies with healthy participants in which contingency awareness was manipulated using a masking procedure in a differential fear conditioning paradigm (for a detailed list of our inclusion criteria see Section 2.2). Particularly, based on the aforementioned problems of the available evidence for contingency unaware fear conditioning, the aim of this review is threefold. First, we will assess the extent to which articles on this topic are affected by the methodological problems described above (i.e., poor masking methodology, inadequate awareness measures, researcher degrees of freedom, and trial-order effects). Second, using meta-analytical tools (i.e., moderator analyses), we will investigate whether there are conditions or measures under which contingency unaware fear conditioning can be consistently observed. Finally, we will investigate whether there is evidence for publication bias in this literature.

2. Method

For the current meta-analysis, PRISMA-guidelines were followed (Moher et al., 2015). In case of deviations of these guidelines, this is explicitly acknowledged.

2.1. Protocol, registration, and materials availability

The protocol for this systematic review and meta-analysis was not publicly registered in a repository. Relevant datafiles (i.e., overview of the search strategy, extracted information from the studies, and the R script for the meta-analysis) are provided through the Open Science Framework (OSF) through the following link: https://osf.io/dy4ac/

2.2. Literature search and inclusion criteria

Relevant articles were identified by a systematic search on three digital databases (PubMed, PsycINFO, and Web of Science[3]; conducted January 13th, 2019), by including relevant articles from a recent systematic review on a closely related topic (i.e., on peripheral physiological responses towards subliminally presented negative affective stimuli, including subliminal CS presentations; van der Ploeg et al., 2017), and by snowballing through the reference lists of relevant recent publications (Schultz and Helmstetter, 2010; Sevenster et al., 2014; Singh et al., 2013). An overview of the results of our search strategy is provided in Fig. 1.

The identified studies were screened according to several selection criteria to determine their inclusion in the meta-analysis. In an initial screening, based on title and abstract conducted by the first author and a graduate student volunteer, studies were selected that used a fear conditioning procedure and a sample of healthy human participants (i.e., no studies with either non-human animals or patients were included). Subsequently, full-texts of the relevant articles were screened for further inclusion in the systematic search and meta-analysis. Particularly, using the participants, interventions, comparisons, and outcomes (PICO) framework (Huang et al., 2006), our inclusion criteria were as follows:

(1) Participants: Only studies on healthy human participants were included.
(2) Interventions: Studies had to use a fear conditioning procedure (i.e., a procedure in which a CS and US are paired and using an aversive stimulus as the US). Studies on eyeblink conditioning, evaluative conditioning, and other classical conditioning procedures were excluded, even though the role of awareness is contested with these classical conditioning procedures as well (see Bar-Anan et al., 2010; Corneille and Stahl, 2018; Weidemann et al., 2016). Furthermore, we only included studies that attempted to mask the presentation of the CSs or the CS-US relationship during the acquisition phase. That is, several studies focused on whether conditioned fear can be expressed unconsciously, not on whether conditioned fear can be acquired unconsciously. As such, these studies did not employ masking procedures during the initial conditioning phase, but only during a subsequent test phase to assess unconscious expression of conditioned fear (e.g., Öhman and Soares, 1993). Because the focus of our meta-analysis is on whether conditioned fear can be *acquired* in the absence of awareness, these studies were excluded.
(3) Comparison group: Only studies that used a within-subjects differential fear conditioning paradigm (i.e., using a CS paired with the US, CS +, and a CS not paired with the US, CS-) were included.
(4) Outcomes: We only included studies that focused specifically on fear conditioning. Particularly, studies had to include a physiological outcome typically measured in fear conditioning studies (e.g., skin conductance responses, fear potentiated startle, heart rate). One exception to this were the studies by Raes and colleagues (Raes and De Raedt, 2011; Raes, De Raedt et al., 2009; Raes, Koster et al., 2010). They used reaction times in a spatial cueing task. Because

---

[3] Search terms: unconscious OR inattentive OR unmindful OR unobservant OR unaware AND fear conditioning. Limited to English language and published after 1970 until now. The searches were executed on January 12th 2019.
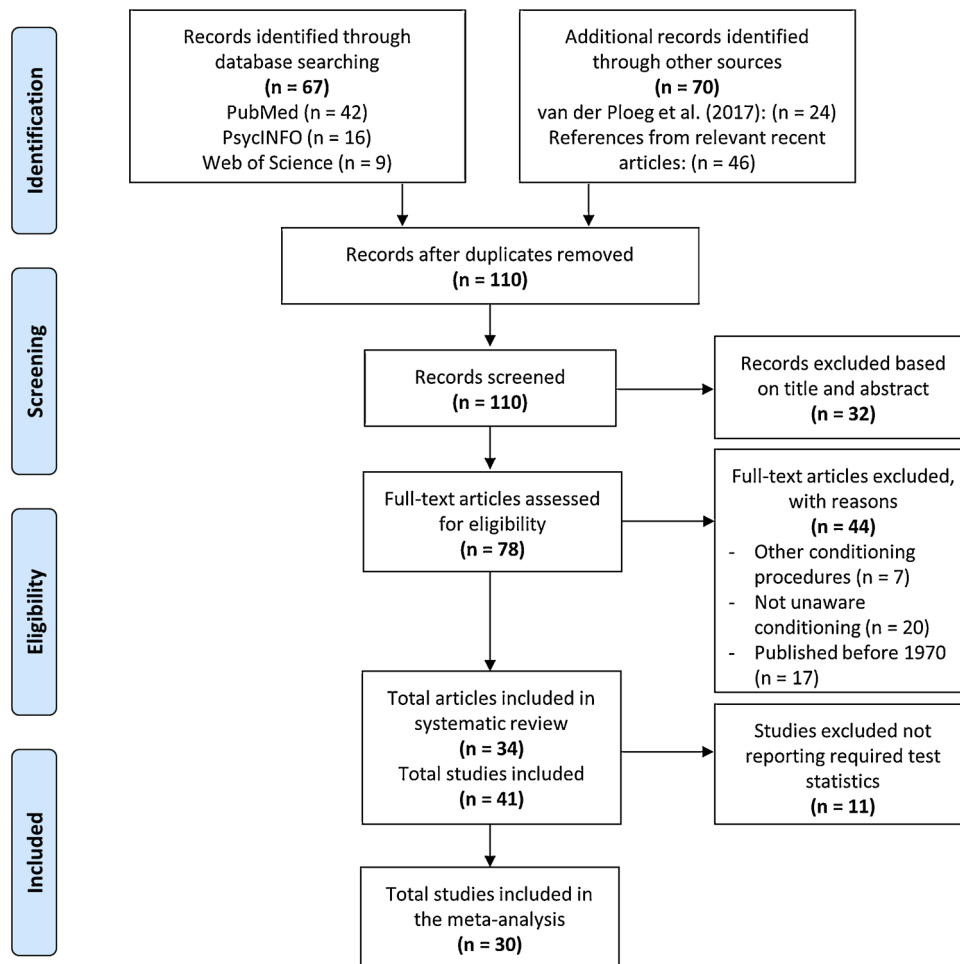
Fig. 1. Flow chart for the literature search.

these studies met all other criteria (i.e., healthy participants, use of a differential fear conditioning procedure, and a masking procedure) and because performance in a reaction time task can be considered to be fairly outside of voluntary control, these studies were also included in the systematic review and meta-analysis.[4]

(5) Time: We only included studies published after 1970 for several reasons. First, there is a lack of standardization with regard to the collection, scoring, and analysis of psychophysiological responses prior to 1970. Especially practices regarding measuring electrodermal activity (i.e., skin conductance) became more standardized following the publication of guidelines (Fowles et al., 1981; Lykken and Venables, 1971). Second, the use of parametric tests and reporting of test statistics is generally lacking in earlier papers. Third, more standardized procedures for assessing awareness were established with the seminal publications by Michael Dawson and colleagues (Dawson, 1970; Dawson and Biferno, 1973; Dawson and Reardon, 1973).

(6) For the meta-analysis and moderator analyses, we only included studies that reported the required test statistics (i.e., the *F*- or *t*-statistic) to calculate the effect size (see Section 2.5). Some studies merely stated whether an effect was significant or not (without the exact *p*-value or test statistics). Because effect size estimates could not be extracted without this information, these studies were excluded from the meta-analysis. However, we did include these

studies into the systematic review to assess the quality of these studies.

We identified a final sample of *k* = 41 different studies (retrieved from 34 different published articles).

2.3. Data extraction and coding

We screened the selected articles for the crucial statistical test demonstrating fear conditioning without awareness. Usually this was a *t*-test or a *F*-test comparing fear responses to the CS + and CS- at the end of a fear conditioning phase for participants who did not demonstrate awareness on the awareness tests. Alternatively, the comparison was between CS + and CS- trials that were not detected by participants. The extracted test statistics were further transformed to obtain the crucial effect size estimate (see Section 2.5).

Furthermore, we coded the identified articles on the following three features. First, we coded the outcome measure. Most of the articles reported several outcome measures. In this case, we selected the outcome measure that was, according to the article, the most sensitive to unaware fear conditioning. We selected only one outcome measure because datapoints in meta-analyses have to be independent and because commonly one outcome measure was identified in the articles as being more sensitive to unaware fear conditioning than the other outcome measures. Selecting one effect size per study is accepted practice in meta-analyses (Quintana, 2015), and it is especially justified in this context given that it may be theoretically argued that certain outcome variables may be insensitive to contingency unaware fear conditioning

---

[4] The results of our systematic review and meta-analysis (see below) were comparable when the studies by Raes and colleagues were excluded from the dataset.

Table 1
Overview of the different studies included in the meta-analysis.

Study	Masking Procedure	Conditioning Procedure	Type of Stimuli	Type of Awareness check	Unaware N/Original N	Outcome measures	Conclusion: Unaware fear conditioning?
Biferno and Dawson (1977)	Tones + distractor tones	Delay	CSs: Tones (800 and 1200 Hz) US: White noise (87 dB)	US expectancy rating	24/60	SCRs	No
Bunce (1999)	Tachistoscopic presentation (2 ms)	Trace	CSs: Unpleasant face (CS+) and pleasant face (CS-) US: Electric shock	VDT and PEQ	8/8	fEMG	Yes
Cornwell et al. (2007)	Dichoptic configured images	Delay	CSs: Colored signs ("+" and "x") US: Electric Shock	VDT and PEQ	28/28 (discrimination sensitivity was included as a continuous predictor)	SCRs	No
Dawson (1970) Experiment 1	Tones + other tones to mask the CS-US relationship	Trace	CSs: Tones (800 and 1200 Hz) US: Electric shock	PEQ	12/20 (non-instructed condition)	SCRs	No
Dawson (1970) Experiment 2	Tones + other tones to mask the CS-US relationship	Trace	CSs: Tones (800 and 1200 Hz) US: Electric shock	PEQ	10/15	SCRs	No
Dawson and Biferno (1973)	Tone CSs (8 s) paired with an electric shock US	Delay	CSs: Tones (800 and 1200 Hz) US: Electric shock	US expectancy rating and PEQ	10/14 (uninstructed group)	SCRs	No
Dawson et al. (1979)	Compound conditioning. Participants partly informed about contingencies	Delay	CSs: Colored lights and tones (750 – 950 Hz) US: Electric shock	US expectancy rating and PEQ	32/43 (partially informed group)	SCRs HR Digital pulse volume	No
Dawson et al. (1986)	Distractor task	Delay	CSs: Fear-relevant or fear-irrelevant pictures (between-subjects) US: Electric shock	Continuous US expectancy rating	28/72 (fear-relevant CSs)	SCRs	No
Dawson et al. (2007)	Picture-picture evaluative conditioning	Delay	CSs: Neutral faces US: IAPS pictures	US expectancy rating and PEQ	37/72 (fear-irrelevant CSs) 29/55	SCRs FPS Affective priming CS ratings	No
Dawson and Reardon (1973) Experiment 1	Instructed vs uninstructed fear conditioning	Trace	CSs: Tones (800 and 1200 Hz) US: Electric shock	Different PEQ (short recall, long recall, short recognition, long recognition)	12/40 (contingent groups)	SCRs	No
Dawson and Reardon (1973) Experiment 2	Uninstructed fear conditioning	Trace	CSs: Tones (800 and 1200 Hz): US: Electric shock	Different PEQ (short recall, short recognition)	28/60	SCRs	No
Esteves et al. (1994) Experiment 1	Backward masking	Trace	CSs: Angry male face (CS+) and happy male face (CS-) US: Electric Shock	VDT (in pilot study)	20/20 (preattentive conditioning group)	SCRs	Yes
Esteves et al. (1994) Experiment 2	Backward masking	Trace	CSs: Angry and happy male face (assignment CS+/CS-counterbalanced) US: Electric shock	VDT (in pilot study)	20/20 (masked angry CS+ group)	SCRs	Yes (only with angry male face as CS+)
Grillon (2002)	Standard conditioning (no masking procedure)	Delay	CSs: Pictures of a flower and mushroom US: Electric shock	PEQ	53/125	SCRs FPS Subjective anxiety	No
Hamm and Vaitl (1996)	No particular masking procedure (CSs presented for 8 s supraliminally)	Delay	CSs: Neutral, negative or erotic IAPS pictures (between-subjects): US: Electric shock	PEQ	16/60 (all different CS type groups)	SCRs FPS HR CS ratings	Yes (with FPS and HR; not SCRs)
Katkin et al. (2001)	Backward masking	Trace	CSs: Pictures of snakes and spiders US: Electric shock	VDT and US expectancy ratings	36/46	SCRs	Yes (also US expectancy; good heartbeat detectors only)

(continued on next page)

Table 1 (*continued*)

Study	Masking Procedure	Conditioning Procedure	Type of Stimuli	Type of Awareness check	Unaware N/Original N	Outcome measures	Conclusion: Unaware fear conditioning?
Klucken et al. (2009)	Picture-picture evaluative conditioning	Delay	CSs: Geometric shapes US: Aversive IAPS pictures	PEQ	18/32	SCRs fMRI activation CS ratings	Yes (fMRI activation [insula] only)
Knight et al. (2003)	Adaptive threshold technique (i.e., intensity CS adjusted according to participants' detection of the CSs)	Delay	CSs: Tones (700 and 1300 Hz) US: White noise (100 dB)	US expectancy ratings	11/11 (analyses compared perceived with unperceived trials)	SCRs	Yes (SCRs; not US expectancy)
Knight et al. (2006)	Adaptive threshold technique	Trace and delay (between-subjects)	CSs: Tones (700 and 1300 Hz) US: White noise (100 dB)	US expectancy ratings	26/26 (see above)	SCRs	Yes (only for SCRs and with delay conditioning)
Knight et al. (2009)	Adaptive threshold technique	Delay	CSs: Tones (700 and 1300 Hz) US: White noise (100 dB)	US expectancy ratings	15/15 (see above)	SCRs fMRI	Yes (SCRs and amygdala activation)
LaBar et al. (2004)	No particular masking procedure (CSs presented for 4 s supraliminally)	Delay	CSs: Colored squares US: White noise (100 – 105 dB)	PEQ	36/87 (all age groups) 7/46 (young adults)	SCRs	Yes (unaware young adults only)
Lipp et al. (2014)	Binocular switch suppression	Delay	CSs: Snake image as CS + and wallaby images as CS- (or vice versa; varied between-subjects) US: Electric shock	PEQ and VDT	18/30 (PEQ) 16/29 (VDT; d' ≤ .5)	SCRs	Yes (both for snake and wallaby CSs)
Marinkovic et al. (1989)	Distractor stimuli to mask the contingency. Half the participants informed about the presence of a contingency	Delay	CSs: Odors (2 CSs and 2 distractors; matched in preference and familiarity) US: Electric shock	US expectancy ratings and PEQ	18/22 (uninstructed group)	SCRs	No
Morris et al. (2001)	Backward masking	Trace	CSs: Angry male faces US: White noise (100 dB)	Target detection task (indicate perception of the target through a button press)	6/6 (analyses compared perceived trial with unperceived trials)	SCRs fMRI activat ion	Yes
Öhman and Soares (1998) Experiment 1	Backward masking (timing parameters based on Öhman and Soares, 1993)	Trace	CSs: Pictures of snakes and spiders, or of mushrooms and flowers (between-subjects) US: Electric shock	None	40/40	SCRs	Yes (only with fear-relevant stimuli)
Öhman and Soares (1998) Experim ent 2	Backward masking, but with a longer CS-US interval for US expectancy ratings	Trace	CSs: Pictures of snakes and spiders US: Electric shock	US expectancy ratings and VDT	10/16 (US expectancy rating group)	SCRs	Yes
Purkis and Lipp (2001)	Distractor task	Delay	CSs: Mountain and lake picture: US Electric shock	US expectancy ratings and PEQ	16/36	SCRs FPS CS ratings	No
Raes et al. (2009)	CSs were cues in a spatial cueing task. Half of the participants were informed about the CS-US contingency	Trace	CSs: Colored quadrangles US: Aversive IAPS picture	US expectancy ratings	12/18 (uninstructed group)	RTs in a spatial cueing task	Yes (though a reversed effect than anticipated; slower RTs with CS + cue)
Raes et al. (2010) Experiment 1	Spatial cueing task (see above) plus backward masking	Trace	CSs: Angry male faces US: Aversive IAPS pictures	PEQ and VDT	32/34 (VDT)	RTs (see above) CS ratings	Yes (RTs only; again reversed; see above)
Raes et al. (2010) Experiment 2	Spatial cueing task (see above) plus backward masking	Trace	CSs: Angry male faces US: White noise (100 dB)	PEQ, VDT, and US expectancy ratings	19/22 (VDT)	RTs (see above) CS ratings	Yes (RTs only; again reversed; see above)
Raes et al. (2010) Experiment 3	Same as Experiment 2 with an additional instructions to improve the sensitivity of the US expectancy ratings	Trace	CSs: Angry male faces US: White noise (104 dB)	PEQ, VDT, and US expectancy ratings	12/22 (US expectancy)	RTs (see above) CS ratings	No
Raes and De Raedt (2011)	Spatial cueing task (see above) plus heartbeat detection task for half of the participants	Trace	CSs: Angry male faces US: White noise (98 dB)	PEQ and VDT	34/34 (VDT performance as a continuous predictor)	RTs (see above)	Yes (only in heartbeat detection task group)
Raio et al. (2012) Experiment 1	Continuous flash suppression	Delay	CSs: Male and female fearful face US: Electric shock	VDT	18/18	SCRs	Yes (early acquisition only)
Raio et al. (2012) Experim ent 2	Continuous flash suppression	Delay	CSs: Male and female fearful face US: Electric shock	VDT	21/24	SCRs	Yes

Table 1 (continued)

Study	Masking Procedure	Conditioning Procedure	Type of Stimuli	Type of Awareness check	Unaware N/Original N	Outcome measures	Conclusion: Unaware fear conditioning?
Schultz and Helmstetter (2010)	Easy or difficult to discriminate sine wave gratings	Delay	CSs: Sine wave gratings; US: Electric shock	US expectancy ratings	8/10 (difficult to discriminate condition)	SCRs	Yes
Sevenster et al. (2014)	Easy or difficult to discriminate sine wave gratings	Delay	CSs: Sine wave gratings; US: Electric shock	US expectancy ratings	18/19 (difficult to discriminate condition)	SCRs FPS	Yes (FPS only)
Singh et al. (2013)	Easy or difficult to discriminate sine wave gratings	Delay	CSs: Sine wave gratings; US: Electric shock	PEQ and US expectancy ratings	26/31 (difficult to discriminate condition)	SCRs	No (effect depends on trial order)
Tabbert et al. (2006)	Distractor task (2-back task). Half the participants were informed about the CS-US contingency	Delay	CSs: Geometric shapes; US: Electric shock	PEQ	17/19 (uninstructed group)	SCRs fMRI activation	Yes (fMRI activation only; amygdala and occipital cortex ROIs)
Tabbert et al. (2011)	See above	Delay	CSs: Geometric shapes	PEQ	41/64 (uninstructed group)	SCRs CS ratings fMRI activation	Yes (fMRI activation only; amygdala, orbital OFC and bilateral hippocampus ROIs)
Weike et al. (2007)	Trace or delay conditioning	Trace and delay (between-subjects)	CSs: Neutral male faces; US: Electric shock	PEQ	8/32 (delay conditioning) 8/32 (trace conditioning)	SCRs FPS	Yes (FPS only and only in the delay conditioning group)
Wiens et al. (2003)	Backward masking	Delay	CSs: Pictures of snakes and spiders; US: Electric shock	PEQ and VDT	65/82	SCRs US expectancy	No (effect depends on trial order)

Notes: CS = conditioned stimulus; US = unconditioned stimulus; VDT = visual discrimination task; PEQ = post-experimental questionnaire; fEMG = facial electromyography; SCRs = skin conductance responses; HR = heart rate; FPS = fear potentiated startle; ROIs = regions of interest.

(Sevenster et al., 2014; Weike et al., 2007). In one article, we were unable to extract the crucial test statistic of the outcome measure of interest (fear potentiated startle) and we selected another outcome measure that was sensitive to unaware fear conditioning instead (heart rate) (Hamm and Vaitl, 1996).

Second, we coded whether the studies used evolutionary fear-relevant or fear-irrelevant CSs. There is some debate about what stimuli precisely constitute evolutionary fear-relevant stimuli. In order to accommodate the claims from most articles and to ensure sufficient statistical power, we coded spider pictures, snake pictures, and angry faces as evolutionary fear-relevant stimuli (Mallan et al., 2013; Öhman, 2009) and geometric figures, grey patterns, sounds, odors (perfumes), and neutral faces as evolutionary fear-irrelevant (see Table 1 for a precise description of the stimuli in the different studies).

Finally, we coded whether studies used a trace or a delay fear conditioning procedure. Trace conditioning procedures were identified as procedures having a temporal gap between CS offset and US onset (i.e., the trace interval). Delay conditioning procedures were defined as studies that presented the US either during CS presentation or immediately at CS offset.

Data extraction and coding was done by the first author.

2.4. Study quality assessment

The identified studies were systematically assessed regarding the extent to which they are affected with the methodological issues described above (i.e., poor masking procedures, poor awareness measures, researcher degrees of freedom, and trial-order effects). Quality assessment was done independently by the first author and a graduate student volunteer using a scoring sheet (see the Supplementary Materials). For each of the four methodological problems, studies were coded as addressing the issue adequately (1) or inadequately (0). Cohen's kappa was calculated on the initial ratings as an indication for the interrater reliability of the quality assessment (Cohen, 1960).

2.5. Meta-analytic procedures

As described previously, the test of interest was whether participants show conditioned fear responses in conditions that prevented or controlled for the awareness of the CS-US contingency. Most typically, this involved a *t*-test or a *F*-test comparing fear responses to the CS + and CS- for unaware participants or undetected trials (compared within the same participants; hence paired sample *t*-tests or repeated measures ANOVAs). Effect sizes were calculated on the basis of this test statistic. Particularly, *F*-statistics were transformed to *t*-statistics ($\sqrt{F}$). Thereafter, Cohen's *d* was calculated using the following formula: Cohen's $d = t/\sqrt{n}$. This way of calculating Cohen's *d* is usually not recommended because it does not allow a direct comparison between between-subjects and within-subjects experiments (i.e., because it takes into account the correlation between the repeated measures in a within-subjects design; Lakens, 2013; Morris and DeShon, 2002). Nevertheless, this was considered adequate here because all studies under consideration and, more generally, nearly all studies within the human fear conditioning literature, usually compare fear responses to the CS + and CS- within participants (for comparable argumentation to use this effect size in the context of human experimental research see Cracco et al., 2018; Hirst et al., 2018). Cohen's *d* was further transformed into Hedges *g* to account for biased estimates of this Cohen's *d* with small sample sizes using the following formula:

$$Hedges\ g = Cohen's\ d * \left[1 - \left(\frac{3}{4*n-1} - 1 \right) \right] \tag{1}$$

The corresponding variance was calculated using the following formula (Cracco et al., 2018; Hirst et al., 2018):

$$V_g = \frac{(n+n)}{(n*n)} + \frac{g^2}{2(n*n)} \qquad (2)$$

The extracted effect sizes were analyzed using the Metafor package in R using a restricted likelihood random-effects model (Viechtbauer, 2010). A random-effects model was used to account for methodological variability between the studies (Hedges and Olkin, 1985; Raudenbush, 2009). Moderator analyses were executed on the extracted effect sizes using either type of outcome measure (skin conductance, other measures), conditioning procedure (trace or delay conditioning), and type of CS (fear-relevant of fear-irrelevant) as a factor. As a measure of consistency of the results of studies included in the meta-analysis we report I^2 (Higgins et al., 2003).

3. Results

3.1. Study characteristics

Table 1 provides a compressed overview of the most important procedural characteristics of the studies included in the meta-analysis. The Supplementary Materials provide a more extensive overview table of the procedural details of the included studies. Out of the 41 identified studies, 16 concluded that unaware fear conditioning cannot take place without contingency awareness, whereas the other 25 studies concluded that unaware fear conditioning can (under some conditions) take place. Overall, sample sizes in the relevant conditions (i.e., under masking conditions) were quite small (median N = 28; range = 6–144) and often a substantial number of participants needed to be excluded because they demonstrated contingency awareness, resulting in even smaller samples (median N = 18; range = 6–65). Skin conductance responses were by far the most commonly used outcome measure (used in 35 studies). The most commonly used measure of contingency awareness was a post-experimental questionnaire (PEQ; 25 studies), followed by US expectancy ratings (18 studies) and visual discrimination tasks (VDT; 14 studies).

3.2. Study quality assessment

The number of studies affected by the four previously described methodological problems is provided in Table 2. As can be seen, a substantial portion of the studies suffered from methodological problems, limiting their interpretability. It may be noted that interrater reliabilities concerning the quality of awareness measures and control of trial-order effects were (near) perfect. However, interrater reliabilities for the assessment of the masking methodology and researcher degrees of freedom were weaker, indicating that assessment of these methodological features was less straightforward. Discrepancies in the assessment (27 out of 164 or 16.46 % of all coded methodological features) were resolved through discussion.

To investigate how the presence of these problems affected the conclusion of the studies (contingency unaware fear conditioning: yes or no), we calculated a sum score of the number of methodological problems accounted for. Hence, a score of four indicates that all described methodological problems were taken into account, whereas a score of zero indicates that none of the methodological problems were

accounted for. Fig. 2 shows the relationship between this sum score and the number of studies providing evidence for contingency unaware fear conditioning. As can be seen, a higher sum score (i.e., higher methodological quality) was related to a lower proportion of studies demonstrating evidence for contingency fear conditioning, $\chi^2(4) = 11.24$, $p = .024$, indicating that reports of contingency unaware fear conditioning are based mostly on methodologically weaker studies.

3.3. Meta-analysis and moderator analyses

Studies for which the effect size could be extracted ($k = 30$) were further analyzed using a random-effects model. The forest plot of the meta-analysis is provided in Fig. 3. The overall meta-analytic effect size was Hedges $g = 0.49$ (95 % CI [0.35, 0.62]), which can be considered medium. The heterogeneity (i.e., percentage of variability in the effect size estimates attributable to heterogeneity among the true effects) was fairly low ($I^2 = 24.64$ %; $Q(29) = 39.96$, $p = 0.085$) (Higgins et al., 2003), indicating fairly low variability between the effects of the different studies. Note, however, that the informational value of this meta-analysis and effect size estimate is limited by the low methodological quality of most studies.

The limited heterogeneity in the meta-analysis indicates that significant moderators of the heterogeneity in effect sizes are unlikely (Viechtbauer, 2010). This was confirmed by moderator analyses. None of the included moderators significantly accounted for the heterogeneity in the meta-analysis (type of outcome measure: $Q(1) = 0.06$, $p = 0.814$; conditioning procedure: $Q(1) = 0.99$, $p = 0.320$; type of CS: $Q(1) = 0.48$, $p = 0.490$). These results indicates that, based on the current set of included studies, there is no supporting evidence for the idea that contingency unaware fear conditioning takes place under a specific set of conditions.

3.4. Publication bias

3.4.1. Egger's regression test

Publication bias was first assessed using Egger's regression test (Egger et al., 1997). This test addresses whether there is a systematic relationship between the size of the observed effects and their standard error (i.e., whether there is an asymmetry in the funnel plot; see Fig. 4). A systematic relationship (i.e., funnel plot asymmetry) is indicative for publication bias. The Egger's test for the included studies was significant, z = 2.84, p = .005. An estimated number of 10 studies were missing on the left side of the funnel plot (see Fig. 4). A trim-and-fill procedure corrected the observed average effect in the meta-analysis to Hedges $g = 0.33$ (95 % CI [0.17, 0.49]). However, the trim-and-fill procedure is considered to be an insufficient procedure to fully correct for publication bias (Carter et al., 2019). That is, it only corrects for publication bias based on observed effect size and not based on whether an effect was significant (see Simonsohn et al., 2014). Therefore, publication bias was further examined using a *p*-curve analysis.

3.4.2. P-curve analysis

The distribution of the significant *p*-values of the included studies was examined using a *p*-curve analysis. *P*-curve analysis was developed by Simonsohn et al. (2014) and provides a way to evaluate the presence

Table 2
Overview of the number of studies affected by the described methodological problems and initial interrater reliability of the assessment.

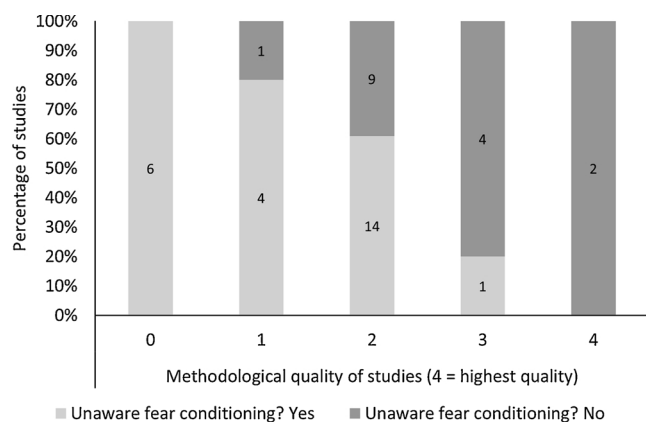| Methodological problem	Number (%) of studies affected	Interrater reliability (Cohen's κ)	Cohen's κ 95 % CI	Approximate significance
Weak masking methodology	17 (41.46 %)	0.42	[0.21, 0.63]	.001
Poor awareness measure	20 (48.78 %)	0.85	[0.70, 1]	< .001
Potential researcher degrees of freedom	15 (36.59 %)	0.40	[0.12, 0.69]	.007
Trial-order effects uncontrolled	38 (92.68 %)	1	[1, 1]	< .001

Fig. 2. Stacked bar chart of the number of studies according to their methodological quality. A higher number indicates better methodological quality (i.e., more methodological issues taken into account). The numbers in the bars reflected the number of individual studies within each category.

of publication bias using exclusively significant results reported in the literature. A *p*-curve analysis plots the distribution of different significant *p*-value levels (i.e., .01, .02, .03, .04, and .05). This *p*-curve follows a known distribution when the null hypothesis is true and with different population effect sizes. When the null hypothesis is true, the distribution of *p*-values is flat, with all different p-values being equiprobable (i.e., the chances of observing a *p*-value of .01, .05, or .99 are identical). When the null-hypothesis is false and there is a certain effect size in the population, smaller *p*-values (i.e., < .025) are more likely than larger *p*-values (i.e., between .025 and .05). When the literature is affected by publication bias, the *p*-curve will deviate from these expected patterns. Particularly, when papers reporting effects with *p*-values smaller than .05 are preferentially reported, *p*-values of these papers will tend to cluster more closely towards "larger" significant *p*-values (i.e., just below the conventional alpha level of .05), particularly when the true population effect size is close to zero.

The results of the *p*-curve analysis are shown in Fig. 5. As input for the p-curve analysis, the test statistics on which the meta-analysis was based (i.e., *t*- or *F*-test statistics) were used (see above and the data files provided through the OSF page associated with this article). A binomial test indicated that there were no more "smaller" (i.e., < .025; n = 9) than "larger" (i.e., > .025; n = 11) significant *p*-values, *p* = .748, indicating a lack of evidential value in the literature on unaware fear conditioning. In contrast, a binomial test assessing the flatness of the curve (i.e., flatter than 33 % power; green line in Fig. 5) indicated statistical significant support for a lack of evidential value (i.e., the observed curve was flatter than the 33 % power curve), *p* = .023. Hence, a *p*-curve analysis indicates that the evidential value for unaware fear conditioning is compromised and is likely affected by publication bias.

4. Discussion

We reviewed studies investigating whether fear conditioning can occur without awareness of the CS-US contingency. The results indicate that the majority of the available studies were affected by methodological problems. In fact, we found that better methodological quality of a study is related to lower odds of reporting evidence for contingency unaware fear conditioning. Furthermore, moderator analyses did not provide evidence for the hypotheses that unaware fear conditioning is more evident for fear measures other than skin conductance responses, is stronger with fear-relevant CSs instead of fear-irrelevant CSs, or is stronger with delay than trace conditioning procedures. Finally, analyses for publication bias revealed evidence for potential bias in the literature on unaware fear conditioning. Particularly, a funnel plot

asymmetry test and a *p*-curve analysis both indicated that the literature on contingency unaware fear conditioning is affected by publication bias. Taken together, these three systematic assessments of the literature indicate that convincing support for the idea that fear conditioning can occur without contingency awareness is currently lacking.

This conclusion may come as a surprise to researchers who are sympathetic to the idea of contingency unaware fear conditioning. For instance, it may be argued that contingency unaware fear conditioning only occurs under some specific conditions and that this meta-analysis with a high number of studies investigating unaware fear conditioning under suboptimal conditions is not informative about those conditions. Furthermore, researchers may argue that there are good theoretical reasons for why contingency unaware fear conditioning should occur. We consider these two arguments here before acknowledging the limitations of our systematic review and summarizing our conclusions.

4.1. Are there conditions under which unaware fear conditioning occurs?

As mentioned, moderator analysis provided little support for the idea that unaware fear conditioning consistently occurs under certain conditions. However, due to the limited number of available studies, we made a coarse classification of the procedural properties and outcome measures of the different studies, potentially obscuring certain findings. Perhaps there was insufficient power to detect effects with the moderator analyses (i.e., an insufficient number of studies investigating unaware fear conditioning under the right circumstances). Though these possibilities cannot be excluded, there are some problems with this reasoning.

First, there is little consistency in the procedures to investigate contingency unaware fear conditioning. The number of different procedures is nearly as large as the number of different labs (see Table 1). If there was indeed one specific set of conditions under which unaware fear conditioning can be consistently observed, it would be expected that different research groups would eventually converge towards using the same procedure. This, however, has not happened even after more than 80 years of research (see the Introduction). Instead, different labs have developed their own procedures to investigate unaware fear conditioning. This argues against the idea that there is one set of specific conditions under which unaware fear conditioning can be consistently observed (see the variety of procedures to mask the CS-US contingency in Table 1).

Second, as indicated in the Introduction, there is a substantial risk that the literature on contingency unaware fear conditioning is filled with false-positive results because fear conditioning studies usually collect many different outcome measures, which can be pre-processed in different ways, and there are different and flexible criteria for the inclusion or exclusion of participants (Lonsdorf et al., 2019; Ney et al., 2018). Multiple ways of analyzing the data allows chance capitalization, which inflates the risk of finding false positive results (Murayama et al., 2014; Simmons et al., 2011). These practices, which have been fairly common in the psychological science literature until recently (Nelson et al., 2018), may have inflated the evidence for contingency unaware fear conditioning and may have introduced spurious (i.e., non-replicable) moderators for the effect. Hence, the proposed conditions under which contingency unaware conditioning occurs are potentially based on unreliably findings. Preregistration and registered reports are required in the research about the conditions under which unaware fear conditioning can occur to ensure that the results cannot be influenced by such flexible statistical decisions and interpretation of the results (Krypotos et al., 2019; van 't Veer and Giner-Sorolla, 2016).

Finally, the proposed moderators for the conditions under which unaware fear conditioning can occur have rarely been tested directly. Often, proposals of moderators are based on the observations of an effect in one study and not in another study (e.g., observing contingency unaware fear conditioning in a study with fear-relevant stimuli but not in a study with fear-irrelevant stimuli). This type of

Author(s) and Year

Hedges' g [95% CI]

Bunce et al 1999	0.58 [-0.44, 1.61]
Cornwell et al 2007	0.29 [-0.26, 0.84]
Dawson et al 1979	0.26 [-0.24, 0.76]
Dawson et al 2007	0.23 [-0.30, 0.76]
Esteves et al 1994 exp1	0.47 [-0.18, 1.12]
Esteves et al 1994 exp2	0.40 [-0.25, 1.04]
Hamm & Vaitl 1996	0.27 [-0.10, 0.64]
Katkin et al 2001	0.40 [-0.09, 0.88]
Klucken et al 2009	1.03 [0.33, 1.72]
Knight et al 2003	0.56 [-0.32, 1.43]
Knight et al 2006	0.48 [-0.32, 1.28]
Knight et al 2009	0.60 [-0.15, 1.35]
LaBar et al 2004	0.83 [-0.27, 1.94]
Lipp et al 2014	0.56 [-0.15, 1.26]
Marinkovic et al 1989	0.05 [-0.58, 0.68]
Öhman & Soares 1998 exp1	1.38 [0.72, 2.05]
Öhman & Soares 1998 exp2	1.77 [0.82, 2.72]
Raes & De Raedt 2011	0.52 [-0.20, 1.25]
Raes et al 2009	0.57 [-0.15, 1.30]
Raes et al 2010 exp1	0.53 [0.02, 1.04]
Raes et al 2010 exp2	0.58 [-0.08, 1.24]
Raes et al 2010 exp3	-0.29 [-1.11, 0.54]
Raio et al 2012 exp1	0.73 [0.05, 1.42]
Raio et al 2012 exp2	0.45 [-0.18, 1.08]
Schultz & Helmstetter 2010	1.30 [0.37, 2.24]
Sevenster et al 2014	0.49 [-0.19, 1.17]
Singh et al 2013	-0.39 [-0.96, 0.17]
Tabbert et al 2006	0.87 [0.16, 1.58]
Weike et al 2007	0.71 [-0.32, 1.74]
Wiens et al 2003	0.40 [-0.13, 0.92]
RE Model	0.49 [0.35, 0.62]

Observed Outcome

Fig. 3. Forest plot of the 30 studies investigating fear conditioning in the absence of CS-US contingency awareness.
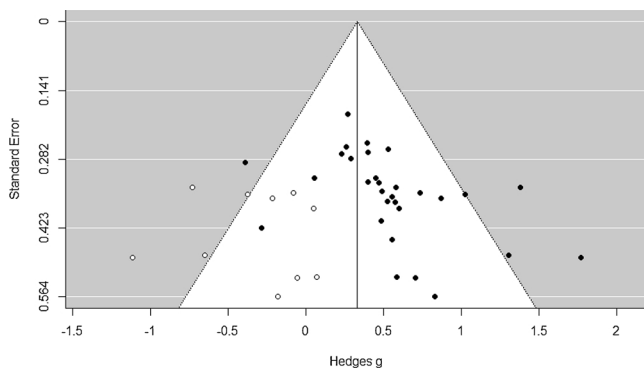
Fig. 4. Funnel plot of the studies included in the meta-analysis. Black dots indicate observed studies and white dots indicate imputed studies correcting for funnel plot asymmetry.

Note: The observed p-curve includes 20 statistically significant (p < .05) results, of which 9 are p < .025. There were 10 additional results entered but excluded from p-curve because they were p > .05.

Fig. 5. P-curve of the significant results demonstrating fear conditioning in the absence of CS-US contingency awareness.

inference is problematic because the difference in statistical significant results (i.e., significant unaware conditioning in one study but not in another study) does not necessarily indicate that the direct comparison of the two conditions would yield statistical significant results (Gelman and Stern, 2006; Nieuwenhuis et al., 2011). Of the reviewed studies, only three studies have tested the effect of stimulus type directly (fear-relevant vs fear-irrelevant) (Esteves, Parra et al., 1994; Lipp et al., 2014; Öhman and Soares, 1998), two have tested the effect of trace vs delay conditioning (Knight et al., 2006; Weike et al., 2007), and none of the studies have directly compared the different outcome measures with each other (such as with a multivariate ANOVA). These numbers of studies indicate that the proposals of moderators for contingency unaware fear conditioning are based on limited empirical evidence. Furthermore, the evidence is not necessarily consistent (e.g., Lipp et al., 2014). Hence, the proposal that contingency unaware fear conditioning occurs under a set of specific conditions requires more and more reliable research.

4.2. Are there theoretical reasons to presuppose contingency unaware fear conditioning?

Some theorists and researchers have argued that contingency unaware fear conditioning is an evolutionary old capacity that has been preserved in humans and demonstrated in non-human animals that most likely do not possess awareness (Öhman and Mineka, 2001; Olsson and Öhman, 2009). As such, contingency unaware fear conditioning in humans is expected on the basis evolutionary continuity (i.e., the preserved capacity to acquire fear based on conditioning experience; which

does not require awareness in non-human animals and thus not in humans either). However, this assumption may need to be reconsidered. That is, within the animal cognition literature, awareness has been a difficult concept to address. For instance, the famous principle of parsimony (or: Morgan's Canon) in animal cognition states that "*In no case is an animal activity to be interpreted as the outcome of the exercise of a higher psychical faculty, if it can be fairly interpreted as the outcome of the exercise of one which stands lower in the psychological scale*" (Morgan, 1903, p. 59). However, this principle has been contested since its conception (Fitzpatrick, 2008), and it has been considerably challenged in recent years by new findings regarding animal cognition. These findings indicate that some animals share more cognitive functions with humans than initially assumed, including basic reasoning, object permanence, and awareness (e.g., Blaisdell et al., 2006; Murphy et al., 2008). Hence, the fact that non-human animals can become fear conditioned is not necessarily an argument in favor of contingency unaware fear conditioning in humans. Non-humans animals may learn in an analogue fashion to humans, indicating the evolutionary continuity of (advanced) cognitive functions supporting learning rather than supporting the possibility of contingency unaware fear conditioning.

Another theoretical argument is based on the organization of the brain. Particularly, fear conditioning has been closely tied to the amygdala, a brain structure that is thought to be part of the "evolutionary old" brain which has few connections with evolutionary more recent brain structures such as the prefrontal cortex. Because higher cognitive functions, such as awareness, are tied to the evolutionary more recent brain structures, the argument has been made that fear conditioning must take place without the cognitive functions served by these brain structures (e.g., Öhman and Mineka, 2001; Tamietto and de Gelder, 2010). However, this neurocognitive model of fear conditioning has also been substantially revised in recent years. Particularly, it has been shown that there are substantial connections between the amygdala and prefrontal brain regions (Stein et al., 2007) and interventions relying on prefrontal regions, such as verbal instructions, appear to modulate defensive reflexes previously related to amygdala activity (e.g., fear potentiated startle; Mertens and De Houwer, 2016) and amygdala activation directly (Phelps et al., 2001). Furthermore, multiple brain areas are implicated in fear conditioning besides the amygdala (Fullana et al., 2016). Hence, the brain model in which fear conditioning is supported specifically by the amygdala and takes place independently of higher cortical regions is most likely too simplified (Pessoa and Adolphs, 2010).

Finally, whether or not awareness is necessary for fear conditioning depends on the exact definition one uses for awareness. According to deterministic views, human behavior is a direct result of the combination of genetic and environmental factors, and awareness is only an epiphenomenon or consequence (but not cause) of the unaware processing of this information. This is a philosophical argument, which is difficult to definitely (dis)prove (Brass et al., 2019; Locke, 1995). However, in human cognition, tremendous progress has been made regarding the research on awareness, indicating that it serves important functions for rapid and flexible adaptation to the environment (Dehaene, 2001; Desender and Van den Bussche, 2012). Such findings make the idea that awareness is only an epiphenomenon less likely. Consequently, the idea that awareness plays an important role in Pavlovian conditioning cannot simply be dismissed on the basis of deterministic theoretical views, and in fact corresponds with the developing insights that awareness supports important cognitive functions. Furthermore, even if awareness is entirely epiphenomenal, it could still be a reliable indicator of the underlying causal neural process(es).

4.3. Limitations of the systematic review and meta-analysis

Several limitations of this review should be noted. First, for a large number of studies (11 out of 41), the required test statistics for calculating effect sizes were not available and were therefore not included in

the overall meta-analysis, moderator analyses, and publication bias analyses. It is not known whether the exclusion of these studies introduces a bias in the outcomes of these analyses.

Second, data extraction and study coding was done solely by the first author. This lack of independent coding may compromise the quality and reliability of the data extraction and coding. Note however that the extracted information is made available through the OSF (https://osf.io/dy4ac/). The extracted information may be checked by interested readers there.

Third, we did not attempt to include unpublished studies in this meta-analysis. This was not attempted because some of the most prominent and productive authors in this research area (e.g., Arne Öhman, Michael Dawson) are no longer actively involved in research, which would have limited the number of unpublished studies that could be obtained. The impact of potential publication bias was partly addressed by the use of a funnel plot asymmetry test and *p*-curve analysis. Furthermore, our systematic review and meta-analysis did not include data from regular fear conditioning studies, in which also often a substantial portion of the sample fails to discover the CS-US contingency. These studies could provide additional information regarding the role of contingency awareness in fear conditioning and the conditions under which contingency unaware fear conditioning may occur. A drawback of regular fear conditioning studies is, however, that they typically do not attempt to prevent the development of contingency awareness and the included measures of contingency awareness are not very elaborated, thereby limiting their methodological quality and informational value regarding contingency unaware fear conditioning (see above).

Fourth, the protocol of this meta-analysis was not preregistered on a public repository (e.g., Prospero). Importantly, the quality assessment of the studies was added during the review process of this paper. As such, the extracted effect sizes and conclusions of the studies were already known to the first assessor during the assessment of the quality of the studies. This could have potentially biased the results. However, the risk of bias was partially mitigated by having the quality assessment done by two independent raters and initial inter-rater reliabilities ranged between fair and excellent (Fleiss et al., 2003), which renders systematic rater biases less plausible.

Fifth and final, this systematic review specifically focused on whether fear conditioning can take place without contingency awareness. Therefore, it does not provide information on whether other types of classical conditioning (e.g., evaluative conditioning; eyeblink conditioning) and other processes (e.g., perception; habituation; sensitization) take place without awareness. Furthermore, awareness is one feature of automaticity of cognitive processes. This systematic review and meta-analysis does not address whether other attributes of automaticity (e.g., involuntary, capacity-free, controllability; see Bargh, 1994; McNally, 1995; Moors and De Houwer, 2006) apply to fear conditioning.

4.4. Conclusions

The evidence for unaware fear conditioning was evaluated. A systematic review of the available studies indicated that many studies are affected by methodological problems, and that the presence of such methodological problems is negatively related to the strength of evidence in favor of contingency unaware fear conditioning. Furthermore, moderator analyses indicated no evidence for conditions under which contingency unaware fear conditioning can be reliably observed. Finally, two tests for publication bias indicated evidence that the literature on contingency unaware fear conditioning is affected by the preferential reporting of positive results. Given these methodological, theoretical, and literature-wide problems, we conclude that there is currently no convincing evidence for contingency unaware fear conditioning in humans.

Author notes

This work was supported by a VICI grant (453-15-005) awarded to Iris Engelhard by the Netherlands Organization for Scientific Research. The funders had no role in the planning, execution, or write-up of this work. We would like to thank Yannick Boddez, Jan De Houwer, Surya Gayet, Miguel Vadillo, and four reviewers for their helpful comments on earlier versions of this paper, Mario Arturo and Rodolfo Bernal for their help with getting access to the full text of all articles included in this review, Ayca Basci with her help in the systematic search and screening of the studies, and Vanessa C. Danzer for her help in coding the methodological quality of the included studies.

Data and material availability statement

Details regarding the systematic search, data extraction and the meta-analysis syntax are available at https://osf.io/dy4ac/.

Declaration of Competing Interest

The authors declare no conflicts of interest with respect to the authorship or the publication of this article.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.neubiorev.2019.11.012.

References[5]

Amodio, D.M., 2018. Social cognition 2.0: an interactive memory systems account. Trends Cogn. Sci. 23 (1), 21–33. https://doi.org/10.1016/j.tics.2018.10.002.

Bakker, M., van Dijk, A., Wicherts, J.M., 2012. The rules of the game called psychological science. Perspect. Psychol. Sci. 7 (6), 543–554.

Bar-Anan, Y., De Houwer, J., Nosek, B., 2010. Evaluative conditioning and conscious knowledge of contingencies: a correlational investigation with large samples. Q. J. Exp. Psychol. 63 (12), 2313–2335. https://doi.org/10.1080/17470211003802442.

Bargh, J.A., 1994. The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In: Wyer, R.S., Srull, T.K. (Eds.), Handbook of Social Cognition. Erlbaum, Hillsdale, NJ, pp. 1–40.

Biferno, M.A., Dawson, M.E., 1977. The onset of contingency awareness and electrodermal classical conditioning: an analysis of temporal relationships during acquisition and extinction. Psychophysiology 14 (2), 164–171. https://doi.org/10.1111/j.1469-8986.1977.tb03370.x.

Blaisdell, A.P., Sawa, K., Leising, K.J., Waldmann, M.R., 2006. Causal reasoning in rats. Science 311 (5763), 1020–1022. https://doi.org/10.1126/science.1121872.

Brass, M., Furstenberg, A., Mele, A.R., 2019. Why neuroscience does not disprove free will. Neurosci. Biobehav. Rev. 102 (January), 251–263. https://doi.org/10.1016/j.neubiorev.2019.04.024.

Brewer, W.F., 1974. There is no convincing evidence for operant or classical conditioning in adult humans. Cognit. Symb. Processes 1–42.

Bunce, S., 1999. Further evidence for unconscious learning: preliminary support for the conditioning of facial EMG to subliminal stimuli. J. Psychiatr. Res. 33 (4), 341–347. https://doi.org/10.1016/S0022-3956(99)00003-5.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14 (May), 365–376. https://doi.org/10.1038/nrn3475.

Carter, E.C., Schönbrodt, F.D., Gervais, W.M., Hilgard, J., 2019. Correcting for bias in psychology: a comparison of meta-analytic methods. Adv. Methods Pract. Psychol. Sci. 2 (2), 115–144. https://doi.org/10.1177/2515245919847196.

Chatterjee, B.B., Eriksen, C.W., 1960. Conditioning and generalization of GSR as a function of awareness. J. Abnorm. Soc. Psychol. 60 (3), 396–403. https://doi.org/10.1037/h0040022.

Chatterjee, B.B., Eriksen, C.W., 1962. Cognitive factors in heart rate conditioning. J. Exp. Psychol. 64 (3), 272–279. https://doi.org/10.1037/h0046192.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46. https://doi.org/10.1177/001316446002000104.

Corneille, O., Stahl, C., 2019. Associative attitude learning: a closer look at evidence and how it relates to attitude models. Personal. Soc. Psychol. Rev. 23 (2). https://doi.org/10.1177/1088868318763261. 108886831876326.

Cornwell, B.R., Echiverri, A.M., Grillon, C., 2007. Sensitivity to masked conditioned stimuli predicts conditioned response magnitude under masked conditions. Psychophysiology 44 (3), 403–406. https://doi.org/10.1111/j.1469-8986.2007.00519.x.

Coursol, A., Wagner, E.E., 1986. Effect of positive findings on submission and acceptance rates: a note on meta-analysis bias. Prof. Psychol.: Res. Pract. 17 (2), 136–137. https://doi.org/10.1037/0735-7028.17.2.136.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., et al., 2018. Automatic imitation: a meta-analysis. Psychol. Bull. 144 (5), 453–500. https://doi.org/10.1037/bul0000143.

Damasio, A., 1994. Descartes' error: Emotion, Reason and the Human Brain. Putnam Publishing., New York.

Dawson, M.E., 1970. Cognition and conditioning: effects of masking the CS-UCS contingency on human GSR classical conditioning. J. Exp. Psychol. 85 (3), 389–396. https://doi.org/10.1037/h0029715.

Dawson, M.E., Biferno, M.a., 1973. Concurrent measurement of awareness and electrodermal classical conditioning. J. Exp. Psychol. 101 (1), 55–62.

Dawson, M.E., Catania, J.J., Schell, A.M., Grings, W.W., 1979. Autonomic classical conditioning as a function of awareness of stimulus contingencies. Biol. Psychol. 9 (1), 23–40. https://doi.org/10.1016/0301-0511(79)90020-6.

Dawson, M.E., Furedy, J.J., 1976. The role of awareness in human differential autonomic classical conditioning: the necessary-gate hypothesis. Psychophysiology 13 (1), 50–53. https://doi.org/10.1111/j.1469-8986.1976.tb03336.x.

Dawson, M.E., Reardon, P., 1973. Construct validity of recall and recognition postconditioning measures of awareness. J. Exp. Psychol. 98 (2), 308–315.

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 (4), 755–766. https://doi.org/10.1037/1528-3542.7.4.755.

Dawson, M.E., Schell, A.M., Banis, H.T., 1986. Greater resistance to extinction of electrodermal responses conditioned to potentially phobic CSs: a noncognitive process? Psychophysiology 23 (5), 552–561. https://doi.org/10.1111/j.1469-8986.1986.tb00673.x.

De Groot, A.D., 2014. The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas]. Acta Psychol. (Amst) 148, 188–194. https://doi.org/10.1016/j.actpsy.2014.02.001.

Dehaene, S., 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition 79 (1–2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2.

Desender, K., Van den Bussche, E., 2012. Is consciousness necessary for conflict adaptation? A state of the art. Front. Hum. Neurosci. 6 (February), 1–13. https://doi.org/10.3389/fnhum.2012.00003.

Diven, K., 1937. Certain determinants in the conditioning of anxiety reactions. J. Psychol. 3 (1), 291–308. https://doi.org/10.1080/00223980.1937.9917499.

Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315 (7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.

Esteves, F., Parra, C., Dimberg, U., Öhman, A., 1994. Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. Psychophysiology 31 (4), 375–385. https://doi.org/10.1111/j.1469-8986.1994.tb02446.x.

Faivre, N., Berthet, V., Kouider, S., 2014. Sustained invisibility through crowding and continuous flash suppression: a comparative review. Front. Psychol. 5 (MAY), 1–13. https://doi.org/10.3389/fpsyg.2014.00475.

Fitzpatrick, S., 2008. Doing away with Morgan's Canon. Mind Lang. 23 (2), 224–246. https://doi.org/10.1111/j.1468-0017.2007.00338.x.

Fleiss, J.L., Levin, B., Paik, M.C., 2003. The measurement of interrater agreement. In: Fleiss, J.L., Levin, B., Paik, M.C. (Eds.), Statistical Methods for Rates and Proportions, 3rd ed. John Wiley, New York, pp. 598–626.

Forstmeier, W., Wagenmakers, E.-J., Parker, T.H., 2017. Detecting and avoiding likely false-positive findings - a practical guide. Biol. Rev. 92 (4), 1941–1968. https://doi.org/10.1111/brv.12315.

Fowles, D.C., Christie, M.J., Edelberg, R., Grings, W.W., Lykken, D.T., Venables, P.H., 1981. Publication recommendations for electrodermal measurements. Psychophysiology 18 (3), 232–239. https://doi.org/10.1111/j.1469-8986.1981.tb03024.x.

Fuhrer, M.J., Baer, P.E., 1965. Differential classical conditioning: verbalization of stimulus contingencies. Science 150 (3705), 1796. https://doi.org/10.1126/science.150.3705.1796-a.

Fuhrer, M.J., Baer, P.E., 1969. Cognitive processes in differential GSR conditioning: effects of a masking task. Am. J. Psychol. 82 (2), 168. https://doi.org/10.2307/1421240.

Fullana, M.A., Harrison, B.J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., Radua, J., 2016. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. Mol. Psychiatry 21 (4), 500–508. https://doi.org/10.1038/mp.2015.88.

Gayet, S., Stein, T., Peelen, M.V., 2019. The danger of interpreting detection differences between image categories: a brief comment on "Mind the snake: fear detection relies on low spatial frequencies" (Gomes, Soares, Silva, & Silva, 2018). Emotion 19 (5), 928–932. https://doi.org/10.1037/emo0000550.

Gelbard-Sagiv, H., Faivre, N., Mudrik, L., Koch, C., 2016. Low-level awareness accompanies "unconscious" high-level processing during continuous flash suppression. J. Vis. 16 (1), 3. https://doi.org/10.1167/16.1.3.

Gelman, A., Loken, E., 2013. The Garden of Forking Paths: Why Multiple Comparisons Can be a Problem, Even When There is No "fishing Expedition" or "p-Hacking" and

---

[5] Articles included in the systematic review.

the Research Hypothesis was Posited Ahead of Time. Department of Statistics, Columbia University.

Gelman, A., Stern, H., 2006. The difference between "significant" and "not significant" is not itself statistically significant. Am. Stat. 60 (4), 328–331. https://doi.org/10.1198/000313006X152649.

Grillon, C., 2002. Associative learning deficits increase symptoms of anxiety in humans. Biol. Psychiatry 51 (11), 851–858. https://doi.org/10.1016/S0006-3223(01)01370-1.

Grings, W.W., 1973. Cognitive factors in electrodermal conditioning. Psychol. Bull. 79 (3), 200–210. https://doi.org/10.1037/h0033883.

Haggard, E.A., 1943. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. J. Exp. Psychol. 33 (4), 257–284. https://doi.org/10.1037/h0058343.

Hamm, A.O., Vaitl, D., 1996. Affective learning: awareness and aversion. Psychophysiology 33 (6), 698–710. https://doi.org/10.1111/j.1469-8986.1996.tb02366.x.

Hamm, A.O., Weike, A.I., 2005. The neuropsychology of fear learning and fear regulation. Int. J. Psychophysiol. 57 (1), 5–14. https://doi.org/10.1016/j.ijpsycho.2005.01.006.

Hedges, L.V., Olkin, I., 1985. Statistical Methods for Meta-Analysis. Academic Press, San Diego, CA.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327 (7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557.

Hirst, R.J., Cragg, L., Allen, H.A., 2018. Vision dominates audition in adults but not children: a meta-analysis of the Colavita effect. Neurosci. Biobehav. Rev. 94 (July), 286–301. https://doi.org/10.1016/j.neubiorev.2018.07.012.

Hofmann, S.G., 2008. Cognitive processes during fear acquisition and extinction in animals and humans: implications for exposure therapy of anxiety disorders. Clin. Psychol. Rev. 28 (2), 199–210. https://doi.org/10.1016/j.cpr.2007.04.009.

Huang, X., Lin, J., Demner-Fushman, D., 2006. Evaluation of PICO as a knowledge representation for clinical questions. AMIA Annual Symposium Proceedings Archive 359–363. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/17238363%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1839740.

Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med. 2 (8), e124. https://doi.org/10.1371/journal.pmed.0020124.

James, W., 1884. What is an emotion? Mind 9 (34), 188–205.

Katkin, E.S., Wiens, S., Öhman, A., 2001. Nonconscious fear conditioning, visceral perception, and the development of gut feelings. Psychol. Sci. 12 (5), 366–370. https://doi.org/10.1111/1467-9280.00368.

Kerr, N.L., 1998. HARKing: hypothesizing after the results are known. Personal. Soc. Psychol. Rev. 2 (3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4.

Klein, R.A., Ratliff, K.A., Vianello, M., Adams, R.B., Bahník, Š., Bernstein, M.J., et al., 2014. Investigating variation in replicability. Soc. Psychol. 45 (3), 142–152. https://doi.org/10.1027/1864-9335/a000178.

Klucken, T., Kagerer, S., Schweckendiek, J., Tabbert, K., Vaitl, D., Stark, R., 2009. Neural, electrodermal and behavioral response patterns in contingency aware and unaware subjects during a picture–picture conditioning paradigm. Neuroscience 158 (2), 721–731. https://doi.org/10.1016/j.neuroscience.2008.09.049.

Knight, D.C., Nguyen, H.T., Bandettini, P.A., 2003. Expression of conditional fear with and without awareness. Proc. Natl. Acad. Sci. U. S. A. 100 (25), 15280–15283. https://doi.org/10.1073/pnas.2535780100.

Knight, D.C., Nguyen, H.T., Bandettini, P.A., 2006. The role of awareness in delay and trace fear conditioning in humans. Cogn. Affect. Behav. Neurosci. 6 (2), 157–162. https://doi.org/10.3758/CABN.6.2.157.

Knight, D.C., Waters, N.S., Bandettini, P.A., 2009. Neural substrates of explicit and implicit fear memory. NeuroImage 45 (1), 208–214. https://doi.org/10.1016/j.neuroimage.2008.11.015.

Krypotos, A.-M., Klugkist, I., Mertens, G., Engelhard, I.M., 2019. A step-by-step guide on preregistration and effective data sharing for psychopathology research. J. Abnorm. Psychol. 128 (6), 517–527. https://doi.org/10.1037/abn0000424.

LaBar, K.S., Cook, C.A., Torpey, D.C., Welsh-Bohmer, K.A., 2004. Impact of healthy aging on awareness and fear conditioning. Behav. Neurosci. 118 (5), 905–915. https://doi.org/10.1037/0735-7044.118.5.905.

Lacey, J.I., Smith, R.L., 1954. Conditioning and generalization of unconscious anxiety. Science 120 (3130), 1045–1052.

Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. 4 (November), 1–12. https://doi.org/10.3389/fpsyg.2013.00863.

Lazarus, R.S., 1982. Thoughts on the relations between emotion and cognition. Am. Psychol. 37 (9), 1019–1024. https://doi.org/10.1037/0003-066X.37.9.1019.

LeDoux, J.E., 2014. Coming to terms with fear. Proc. Natl. Acad. Sci. U. S. A. 111 (8), 2871–2878. https://doi.org/10.1073/pnas.1400335111.

Lee, C.J., Sugimoto, C.R., Zhang, G., Cronin, B., 2013. Bias in peer review. J. Am. Soc. Inf. Sci. Technol. 64 (1), 2–17. https://doi.org/10.1002/asi.22784.

Levine, T.R., Asada, K.J., Carpenter, C., 2009. Sample sizes and effect sizes are negatively correlated in meta-analyses: evidence and implications of a publication bias against nonsignificant findings. Commun. Monogr. 76 (3), 286–302. https://doi.org/10.1080/03637750903074685.

Lipp, O.V., Kempnich, C., Jee, S.H., Arnold, D.H., 2014. Fear conditioning to subliminal fear relevant and non fear relevant stimuli. PLoS One 9 (9), e99332. https://doi.org/10.1371/journal.pone.0099332.

Locke, E.A., 1995. Beyond determinism and materialism, or isn't it time we took consciousness seriously? J. Behav. Ther. Exp. Psychiatry 26 (3), 265–273. https://doi.org/10.1016/0005-7916(95)00026-V.

Lonsdorf, T.B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher,

A., et al., 2019. How to Not Get Lost in the Garden of Forking Paths: Lessons Learned from Human Fear Conditioning Research Regarding Exclusion Criteria. Preprint. https://doi.org/10.31234/osf.io/6m72g.

Lonsdorf, T.B., Menz, M.M., Andreatta, M., Fullana, M.A., Golkar, A., Haaker, J., et al., 2017. Don't fear 'fear conditioning': methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neurosci. Biobehav. Rev. 77, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026.

Lovibond, P.F., Liu, J.C.J., Weidemann, G., Mitchell, C.J., 2011. Awareness is necessary for differential trace and delay eyeblink conditioning in humans. Biol. Psychol. 87 (3), 393–400. https://doi.org/10.1016/j.biopsycho.2011.05.002.

Lovibond, P.F., Shanks, D.R., 2002. The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. J. Exp. Psychol.: Anim. Behav. Processes 28 (1), 3–26. https://doi.org/10.1037//0097-7403.28.1.3.

Lykken, D.T., Venables, P.H., 1971. Direct measurement of skin conductance: a proposal for standardization. Psychophysiology 8 (5), 656–672. https://doi.org/10.1111/j.1469-8986.1971.tb00501.x.

Mallan, K.M., Lipp, O.V., Cochrane, B., 2013. Slithering snakes, angry men and out-group members: what and whom are we evolved to fear? Cogn. Emot. 27 (7), 1168–1180. https://doi.org/10.1080/02699931.2013.778195.

Marinkovic, K., Schell, A.M., Dawson, M.E., 1989. Awareness of the CS-UCS contingency and classical conditioning of skin conductance responses with olfactory CSs. Biol. Psychol. 29 (1), 39–60. https://doi.org/10.1016/0301-0511(89)90049-5.

McAndrew, A., Jones, F.W., McLaren, R.P.L., McLaren, I.P.L., 2012. Dissociating expectancy of shock and changes in skin conductance: an investigation of the Perruchet effect using an electrodermal paradigm. J. Exp. Psychol.: Anim. Behav. Processes 38 (2), 203–208. https://doi.org/10.1037/a0026718.

McNally, R.J., 1987. Preparedness and phobias: a review. Psychol. Bull. 101 (2), 283–303. https://doi.org/10.1037/0033-2909.101.2.283.

McNally, R.J., 1995. Automaticity and the anxiety disorders. Behav. Res. Ther. 33 (7), 747–754. https://doi.org/10.1016/0005-7967(95)00015-P.

Mertens, G., Boddez, Y., Sevenster, D., Engelhard, I.M., De Houwer, J., 2018. A review on the effects of verbal instructions in human fear conditioning: empirical findings, theoretical considerations, and future directions. Biol. Psychol. 137, 49–64. https://doi.org/10.1016/j.biopsycho.2018.07.002.

Mertens, G., De Houwer, J., 2016. Potentiation of the startle reflex is in line with contingency reversal instructions rather than the conditioning history. Biol. Psychol. 113, 91–99. https://doi.org/10.1016/j.biopsycho.2015.11.014.

Mineka, S., Öhman, A., 2002. Phobias and preparedness: the selective, automatic, and encapsulated nature of fear. Biol. Psychiatry 52 (10), 927–937. https://doi.org/10.1016/S0006-3223(02)01669-4.

Mineka, S., Zinbarg, R., 2006. A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. Am. Psychol. 61 (1), 10–26. https://doi.org/10.1037/0003-066X.61.1.10.

Mitchell, C.J., De Houwer, J., Lovibond, P.F., 2009. The propositional nature of human associative learning. Behav. Brain Sci. 32 (2), 183–198. https://doi.org/10.1017/S0140525X09000855. discussion 198-246.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 1–9. https://doi.org/10.1186/2046-4053-4-1.

Moors, A., Boddez, Y., De Houwer, J., 2017. The power of goal-directed processes in the causation of emotional and other actions. Emot. Rev. 9 (4), 310–318. https://doi.org/10.1177/1754073916669595.

Moors, A., De Houwer, J., 2006. Automaticity: a theoretical and conceptual analysis. Psychol. Bull. 132 (2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297.

Morgan, L.C., 1903. An Introduction to Comparative Psychology. The Walter Scott Publishing CO., LTD, London.

Morris, J.S., Buchel, C., Dolan, R.J., 2001. Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. NeuroImage 13 (6), 1044–1052. https://doi.org/10.1006/nimg.2000.0721.

Morris, S.B., DeShon, R.P., 2002. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol. Methods 7 (1), 105–125. https://doi.org/10.1037/1082-989X.7.1.105.

Murayama, K., Pekrun, R., Fiedler, K., 2014. Research practices that can prevent an inflation of false-positive rates. Personal. Soc. Psychol. Rev. 18 (2), 107–118. https://doi.org/10.1177/1088868313496330.

Murphy, R.A., Mondragon, E., Murphy, V.A., 2008. Rule learning by rats. Science 319 (5871), 1849–1851. https://doi.org/10.1126/science.1151564.

Nelson, L.D., Simmons, J., Simonsohn, U., 2018. Psychology's renaissance. Annu. Rev. Psychol. 69 (1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836.

Ney, L.J., Wade, M., Reynolds, A., Zuj, D.V., Dymond, S., Matthews, A., Felmingham, K.L., 2018. Critical evaluation of current data analysis strategies for psychophysiological measures of fear conditioning and extinction in humans. Int. J. Psychophysiol. 134 (October), 95–107. https://doi.org/10.1016/j.ijpsycho.2018.10.010.

Nieuwenhuis, S., Forstmann, B.U., Wagenmakers, E.-J., 2011. Erroneous analyses of interactions in neuroscience: a problem of significance. Nat. Neurosci. 14 (9), 1105–1107. https://doi.org/10.1038/nn.2886.

Nosek, B.A., Spies, J.R., Motyl, M., 2012. Scientific utopia. Perspect. Psychol. Sci. 7 (6), 615–631. https://doi.org/10.1177/1745691612459058.

Novick, M.R., 1966. The axioms and principal results of classical test theory. J. Math. Psychol. 3 (1), 1–18.

Öhman, A., 2009. Of snakes and faces: an evolutionary perspective on the psychology of fear. Scand. J. Psychol. 50 (6), 543–552. https://doi.org/10.1111/j.1467-9450.2009.00784.x.

Öhman, A., Mineka, S., 2001. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. Psychol. Rev. 108 (3), 483–522. https://doi.org/10.

1037//0033-295X.108.3.483.

Öhman, A., Soares, J.J., 1993. On the automatic nature of phobic fear: conditioned electrodermal responses to masked fear-relevant stimuli. J. Abnormal Psychol. 102 (1), 121–132. https://doi.org/10.1037/0021-843X.102.1.121.

Öhman, A., Soares, J.J.F., 1998. Emotional conditioning to masked stimuli: expectancies for aversive outcomes following. J. Exp. Psychol.: Gen. 127 (1), 69.

Olsson, A., Öhman, A., 2009. Is cultivating "biological blindness" a viable route to understanding behavioral phenomena? Behav. Brain Sci. 32 (02), 220. https://doi.org/10.1017/S0140525X09001101.

Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349 (6251), aac4716. https://doi.org/10.1126/science.aac4716.

Pavlov, I.P., 1928. Lectures on Conditioned Reflexes: Twenty-Five Years of Objective Study of the Higher Nervous Activity (Behaviour) of Animals. Liverwright Publishing Corporation, New York. https://doi.org/10.1037/11081-000.

Pessoa, L., Adolphs, R., 2010. Emotion processing and the amygdala: from a "low road" to "many roads" of evaluating biological significance. Nat. Rev. Neurosci. 11 (11), 773–782. https://doi.org/10.1038/nrn2920.

Pessoa, L., Japee, S., Sturman, D., Ungerleider, L.G., 2006. Target visibility and visual awareness modulate amygdala responses to fearful faces. Cereb. Cortex 16 (3), 366–375. https://doi.org/10.1093/cercor/bhi115.

Phelps, E.A., O'Connor, K.J., Gatenby, J.C., Gore, J.C., Grillon, C., Davis, M., 2001. Activation of the left amygdala to a cognitive representation of fear. Nat. Neurosci. 4 (4), 437–441. https://doi.org/10.1038/86110.

Pineles, S.L., Orr, M.R., Orr, S.P., 2009. An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. Psychophysiology 46 (5), 984–995. https://doi.org/10.1111/j.1469-8986.2009.00852.x.

Ploog, B.O., 2012. Classical conditioning. In: Ramachandran, V.S. (Ed.), Encyclopedia of Human Behavior, 2nd ed. Academic Press, Cambridge, Massachusetss, pp. 484–491.

Purkis, H.M., Lipp, O.V., 2001. Does affective learning exist in the absence of contingency awareness? Learn. Motiv. 32 (1), 84–99. https://doi.org/10.1006/lmot.2000.1066.

Quintana, D.S., 2015. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. Front. Psychol. 6 (October), 1–9. https://doi.org/10.3389/fpsyg.2015.01549.

Rachman, S., 1991. Neo-conditioning and the classical theory of fear acquisition. Clin. Psychol. Rev. 11 (2), 155–173. https://doi.org/10.1016/0272-7358(91)90093-A.

Raes, A.K., De Raedt, R., 2011. Interoceptive awareness and unaware fear conditioning: are subliminal conditioning effects influenced by the manipulation of visceral self-perception? Conscious. Cogn. 20 (4), 1393–1402. https://doi.org/10.1016/j.concog.2011.05.009.

Raes, A.K., De Raedt, R., Fias, W., Koster, E.H.W., Van Damme, S., 2009. Does contingency awareness mediate the influence of emotional learning on the cueing of visual attention? Psychol. Res. 73 (1), 107–113. https://doi.org/10.1007/s00426-008-0141-y.

Raes, A.K., Koster, E.H.W., Van Damme, S., Fias, W., De Raedt, R., 2010. Aversive conditioning under conditions of restricted awareness: effects on spatial cueing. Q. J. Exp. Psychol. 63 (12), 2336–2358. https://doi.org/10.1080/17470218.2010.492995.

Raio, C.M., Carmel, D., Carrasco, M., Phelps, E.A., 2012. Nonconscious fear is quickly acquired but swiftly forgotten. Curr. Biol. 22 (12), R477–R479. https://doi.org/10.1016/j.cub.2012.04.023.

Raudenbush, S.W., 2009. Analyzing effect sizes: random effects models. In: Cooper, H., Hedges, L.V., Valentine, J.C. (Eds.), The Handbook of Research Synthesis and Meta-Analysis, 2nd ed. Russell Sage Foundation, New York, pp. 295–315.

Rescorla, R.A., 1988. Pavlovian conditioning. It's not what you think it is. Am. Psychol. 43 (3), 151–160.

Rosenthal, R., 1979. The file drawer problem and tolerance for null results. Psychol. Bull. 86 (3), 638–641.

Schultz, D.H., Helmstetter, F.J., 2010. Classical conditioning of autonomic fear responses is independent of contingency awareness. J. Exp. Psychol.: Anim. Behav. Processes 36 (4), 495–500. https://doi.org/10.1037/a0020263.

Sevenster, D., Beckers, T., Kindt, M., 2014. Fear conditioning of SCR but not the startle reflex requires conscious discrimination of threat and safety. Front. Behav. Neurosci. 8 (32). https://doi.org/10.3389/fnbeh.2014.00032.

Shanks, D.R., 2017. Regressive research: the pitfalls of post hoc data selection in the study of unconscious mental processes. Psychon. Bull. Rev. 24 (3), 752–775. https://doi.org/10.3758/s13423-016-1170-y.

Shanks, D.R., Berry, C.J., 2012. Are there multiple memory systems? Tests of models of

implicit and explicit memory. Q. J. Exp. Psychol. 65 (8), 1449–1474. https://doi.org/10.1080/17470218.2012.691887.

Shanks, D.R., St. John, M.F., 1994. Characteristics of dissociable human learning-systems. Behav. Bain Sci. 17 (3), 367–395. https://doi.org/10.1017/S0140525X00035032.

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22 (11), 1359–1366. https://doi.org/10.1177/0956797611417632.

Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. P-Curve: a key to the file-drawer. J. Exp. Psychol.: Gen. 143 (2), 534–547. https://doi.org/10.1037/a0033242.

Singh, K., Dawson, M.E., Schell, A.M., Courtney, C.G., Payne, A.F.H., 2013. Can human autonomic classical conditioning occur without contingency awareness? The critical importance of the trial sequence. Biol. Psychol. 93 (1), 197–205. https://doi.org/10.1016/j.biopsycho.2013.02.007.

Soares, J.J.F., Öhman, A., 1993. Backward masking and skin conductance responses after conditioning to nonfeared but fear-relevant stimuli in fearful subjects. Psychophysiology 30 (5), 460–466. https://doi.org/10.1111/j.1469-8986.1993.tb02069.x.

Squire, L.R., 2004. Memory systems of the brain: a brief history and current perspective. Neurobiol. Learn. Mem. 82, 171–177. https://doi.org/10.1016/j.nlm.2004.06.005.

Stein, J.L., Wiedholz, L.M., Bassett, D.S., Weinberger, D.R., Zink, C.F., Mattay, V.S., Meyer-Lindenberg, A., 2007. A validated network of effective amygdala connectivity. NeuroImage 36 (3), 736–745. https://doi.org/10.1016/j.neuroimage.2007.03.022.

Tabbert, K., Merz, C.J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O.T., Stark, R., 2011. Influence of contingency awareness on neural, electrodermal and evaluative responses during fear conditioning. Soc. Cogn. Affect. Neurosci. 6 (4), 495–506. https://doi.org/10.1093/scan/nsq070.

Tabbert, K., Stark, R., Kirsch, P., Vaitl, D., 2006. Dissociation of neural responses and skin conductance reactions during fear conditioning with and without awareness of stimulus contingencies. NeuroImage 32 (2), 761–770. https://doi.org/10.1016/j.neuroimage.2006.03.038.

Tamietto, M., de Gelder, B., 2010. Neural bases of the non-conscious perception of emotional signals. Nat. Rev. Neurosci. 11 (10), 697–709. https://doi.org/10.1038/nrn2889.

Vadillo, M.A., Konstantinidis, E., Shanks, D.R., 2016. Underpowered samples, false negatives, and unconscious learning. Psychon. Bull. Rev. 23 (1), 87–102. https://doi.org/10.3758/s13423-015-0892-6.

Vadillo, M.A., Linssen, D., Orgaz, C., Parsons, S., Shanks, D.R., 2019. Unconscious or underpowered? Probabilistic cuing of visual attention. J. Exp. Psychol. Gen. https://doi.org/10.1037/xge0000632.

van 't Veer, A.E., Giner-Sorolla, R., 2016. Pre-registration in social psychology—a discussion and suggested template. J. Exp. Soc. Psychol. 67, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004.

van der Ploeg, M.M., Brosschot, J.F., Versluis, A., Verkuil, B., 2017. Peripheral physiological responses to subliminally presented negative affective stimuli: a systematic review. Biol. Psychol. 129 (June 2016), 131–153. https://doi.org/10.1016/j.biopsycho.2017.08.051.

Vermeiren, A., Cleeremans, A., 2012. The validity of d' measures. PLoS One 7 (2), e31595. https://doi.org/10.1371/journal.pone.0031595.

Vervliet, B., Craske, M.G., Hermans, D., 2013. Fear extinction and relapse: state of the art. Annu. Rev. Clin. Psychol. 9 (1), 215–248. https://doi.org/10.1146/annurev-clinpsy-050212-185542.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3).

Weidemann, G., Satkunarajah, M., Lovibond, P.F., 2016. I think, therefore eyeblink: the importance of contingency awareness in conditioning. Psychol. Sci. 27 (4), 467–475. https://doi.org/10.1177/0956797615625973.

Weike, A.I., Schupp, H.T., Hamm, A.O., 2007. Fear acquisition requires awareness in trace but not delay conditioning. Psychophysiology 44 (1), 170–180. https://doi.org/10.1111/j.1469-8986.2006.00469.x.

Wiens, S., Katkin, E.S., Öhman, A., 2003. Effects of trial order and differential conditioning on acquisition of differential shock expectancy and skin conductance conditioning to masked stimuli. Psychophysiology 40 (6), 989–997. https://doi.org/10.1111/1469-8986.00117.

Yong, E., 2012. Replication studies: bad copy. Nature 485 (7398), 298–300. https://doi.org/10.1038/485298a.

Zajonc, R.B., 1980. Feeling and thinking: preferences need no inferences. Am. Psychol. 35 (2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151.