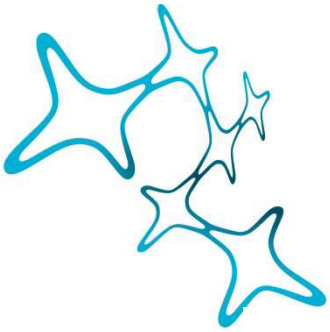
Tracking fear learning with pupillometry: Psychophysiological and neuroimaging investigations

Laura Leuchs



Graduate School of Systemic Neurosciences

LMU Munich

Dissertation der
Graduate School of Systemic Neurosciences
Ludwig-Maximilians-Universität München

15. August, 2018

Supervisor:

Dr. Victor Spoormaker

Project Group Psychophysiology

Max Planck Institute of Psychiatry, Munich

First Reviewer: Dr. Victor Spoormaker

Second Reviewer: Dr. Kathrin Koch

Date of Defense: 23.11.2018

Contents

1. Introduction

1.1 Fear research and the RDoC initiative

Fear is an evolutionary adaptive mechanism: it enables an organism to react to threat by preparing the body for 'fight or flight'. However, the regulation of fear can be altered in a way that it is no longer in proportion to environmental threats. As fear is an aversive state, its overexpression can imply subjective suffering and avoidance behavior and can ultimately lead to impairment in everyday life. Fear-related symptoms are therefore core features of different psychiatric disorders like post-traumatic stress disorder (PTSD), panic disorder or different kinds of phobias (American Psychiatric Association, 2013). Specific symptoms can consist of recurrent panic attacks, excessive fear expression towards stimuli or situations which are not inherently dangerous, avoidance of phobic stimuli, intrusive memories of fearful events, nightmares or sustained states of anxiety.

To determine the presence of a psychiatric disorder, the most commonly employed diagnostic classification systems are the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013) or the International Statistical Manual of Diseases (World Health Organization (WHO), 1992). These classification systems are based on clinical observations and comprise different diagnostic categories with (partly overlapping) lists of symptoms. A diagnosis is given only if a specific amount of symptoms out of a defined cluster is present for a minimum duration and if these symptoms are associated with subjective suffering.

Such a classification of psychiatric disorders is potentially problematic for several reasons. First, the diagnostic categories are not informed by their biological mechanisms, which may imply problems for the correct grouping of patients. On one hand, individuals with overlapping symptom patterns may be assigned the same diagnosis, even though diverging biological mechanisms may be causal for their symptoms (heterogeneity). On the other hand, alterations in one common mechanism may display in different phenotypes (pleitropy), but may be responsive to the same treatment. Such categorization problems might be reflected in high levels of comorbidity, in particular between stress-related disorders such as PTSD, anxiety disorders and major depressive disorder (Flory & Yehuda, 2015; Lang, McTeague, & Bradley, 2016; Rytwinski, Scur, Feeny, & Youngstrom, 2013). Moreover, clinical interviews or self-reports can indicate dysregulated fear, but they don't constitute objective measures of fear processing. Interviews and self-reports are restricted to observable or consciously accessible symptoms. Biological measures instead could capture dysregulations independent of conscious awareness and are inherently more objective than verbal reports. They could furthermore offer quantitative, continuous readouts, which seem more appropriate for describing dysregulations than is given by categorial psychiatric diagnoses.

In the recent years, efforts have been made to improve the mechanistic understanding of mental disorders, by targeting their neurobiological underpinnings. The Research Domain Criteria (RDoC) initiative, started by the National Institute of Mental Health (NIMH), offers a complementary approach for the classification of mental disorders, which cuts across the established categories of current diagnostic systems (Cuthbert, 2014; Insel et al., 2010; Morris & Cuthbert, 2012). This initiative provides a framework to characterize individuals among five major domains on multiple observation levels (ranging from genetics, cell circuitry, and physiology to observable behavior). Each of the major domains relates to a different psychological construct and has been linked to specific neural circuitry (Kozak & Cuthbert, 2016). The five domains are: negative valence, positive valence, cognitive processes, social processes and arousal or regulatory systems, each with several sub-categories. “Fear” represents a sub-construct in the RDoC domain of “negative valence” and can be targeted with a variety of tasks and measures.

1.2 Fear conditioning as experimental and etiological model in psychiatric research

Experimentally, fear is commonly modeled with fear conditioning and extinction tasks. These tasks are not only highly translatable between human and animal models (Milad & Quirk, 2012), they also constitute etiological models for the development and maintenance of fear-related psychiatric disorders. For this reason, they are widely employed in fear research and their use is endorsed for the investigation of fear by the RDoC initiative.

During fear acquisition, a neutral stimulus is repeatedly paired with an aversive event, a so-called unconditioned stimulus (US). With proceeding learning of this association, the neutral stimulus becomes a conditioned stimulus (CS+) which predicts the aversive outcome and can by itself elicit fear responses. In differential fear conditioning, another neutral stimulus is presented but is never paired with the aversive US. With proceeding learning, this stimulus becomes a cue for safety (CS-). During extinction, the former CS+ is presented repetitively without the occurrence of the US. This should lead to the formation of a (context-dependent) extinction memory, which can suppress fear responses to the former CS+ (Bouton & Moody, 2004). At a later time point, the retention of the extinction memory can be assessed by presenting the CS+ again, or it can for example be probed how readily fear is reinstated after the presentation of an unannounced US (for an overview of various conditioning procedures see Lonsdorf et al., 2017). Theoretically, differential fear conditioning leads to discrimination learning between CS+ and CS- and expresses in higher fear responses to CS+ as opposed to CS-. With proceeding extinction, fear responses to CS+ decrease and stimulus discrimination diminishes.

The interpretation of PTSD and anxiety disorders in the scope of fear memory formation and extinction learning can enhance our understanding of their emergence, maintenance and biological foundations. For

translating the experimental model of fear acquisition and extinction to pathological fear, PTSD can serve as a useful example: during an aversive or traumatizing event (reflecting a US), the surrounding context or present stimuli can form strong memory associations with the aversive event. Even though this context and concurrently present stimuli may initially be neutral (CS), they can later serve as reminders or indicators of the associated aversive event and elicit fear responses (i.e. serve as CS+). A strong CS-US association may hinder extinction learning and the attenuation of fear, even if the US is no longer present during repetitive exposure to the conditioned context or stimulus. This in turn can explain the persistence of pathological fear and avoidance of the formerly neutral context in everyday life, as it elicits strong aversive fear responses. Pittig, Treanor, LeBeau, and Craske (2018) offer a detailed description how fear acquisition, impaired extinction learning, avoidance, generalization and the return of fear can contribute to the development and persistence of pathological fear (for PTSD specifically see Briscione, Jovanovic, & Norrholm, 2014; Careaga, Girardi, & Suchecki, 2016; VanElzakker, Dahlgren, Davis, Dubois, & Shin, 2014).

Delineating the learning mechanisms behind pathological fear can improve individualized treatment and therapeutic success (Craske, Hermans, & Vervliet, 2018; Smith, Doran, Sippel, & Harpaz-Rotem, 2017). For example the rationale behind the treatment of pathological fear with exposure therapy is based on extinction learning (Craske et al., 2018; Pittig et al., 2018). During exposure therapy, the patient is repetitively confronted with the avoided stimulus or context (without the US), with the aim of creating an extinction memory. Particularly the impairment of such extinction learning or fear inhibition has been proposed as important mechanism underlying PTSD (Jovanovic & Ressler, 2010; Milad & Quirk, 2012) and anxiety disorders (Craske et al., 2018; Milad & Quirk, 2012). Vervliet, Craske, and Hermans (2013) argue that not only the initial formation of an extinction memory, but especially its consolidation and lasting retention is predictive for long-term therapy success, which makes recall of extinction of fear a particularly interesting subject for psychiatric research.

Numerous studies have used fear acquisition and extinction to investigate differences between psychiatric populations and healthy controls. These tasks have revealed patient-control differences for patients suffering from PTSD (e.g. reviewed by Francati, Vermetten, & Bremner, 2007; Norrholm & Jovanovic, 2018), panic or anxiety disorders (Duits, Cath, Heitland, & Baas, 2016) and yielded initial evidence for obsessive-compulsive disorder (OCD; Geller et al., 2017; Milad et al., 2013). Patient-control differences are thereby not restricted to fear acquisition and extinction, but have also revealed alterations in the recall of extinction memory for PTSD, OCD and panic disorder (Michael, Blechert, Vriends, Margraf, & Wilhelm, 2007; Milad et al., 2013; Milad et al., 2008; Milad et al., 2009).

There have been attempts to summarize the findings from clinical studies addressing different psychiatric populations. Nees, Heinrich, and Flor (2015) have reviewed clinical studies employing fear conditioning and summarize which fear learning processes have yielded patient-control differences for specific psychiatric

disorders (see Figure 1). In a recent meta-analysis, Duits et al. (2015) analyzed data from fear conditioning studies including patient-control comparisons across multiple disorders (PTSD, OCD and disorders of the anxiety spectrum). They found that, on average, patients responded more strongly to CS- during fear acquisition (possibly reflecting less fear suppression in response to safety cues) and showed prolonged stimulus discrimination during extinction learning (i.e. by responding stronger to the extinguished CS+).

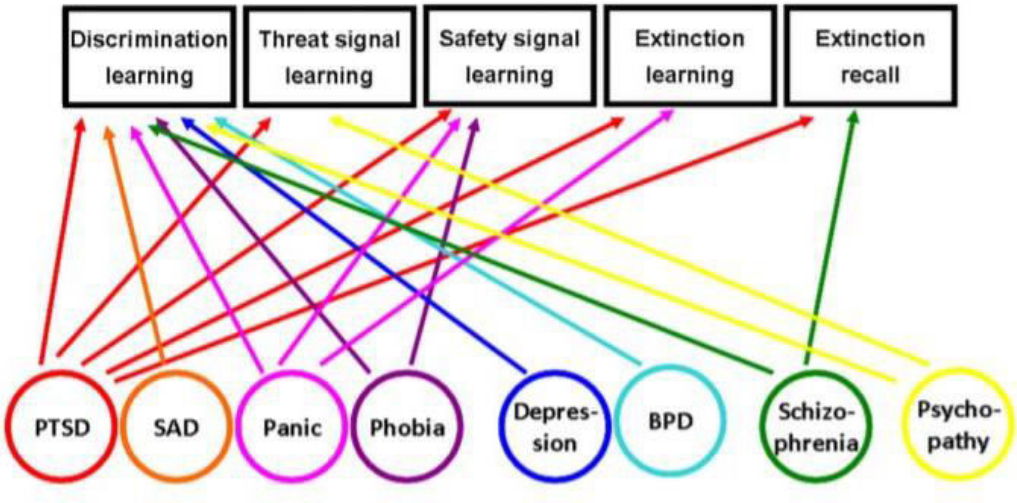


Figure 1. Summary of clinical findings relating psychiatric disorders to alterations in specific mechanisms relevant to fear learning. Arrows indicate which mechanisms were found to be altered in association with different psychiatric diagnoses. The summarized findings stem from studies employing behavioral/subjective, physiological, and/or neural readouts of fear learning.
'Discrimination learning' refers to increased responsiveness to CS+ compared to CS-; 'Threat signal learning' denotes an increased responsiveness to CS+ during fear acquisition, specifically; 'Safety signal learning' refers to an increased responsiveness to the non-paired CS- during fear acquisition; 'Extinction learning' denotes increased responsiveness to CS+ than to CS- during extinction; 'Extinction recall' refers to the occurrence of conditioned responses to previously extinguished CS+. SAD = social anxiety disorder; BPD = bipolar disorder. From Nees et al. (2015), reproduced with kind permission from Elsevier.

1.3 Readouts of conditioned fear: state of the art

To date, most clinical studies define patient and control groups according to the above-mentioned diagnostic systems. This might, as discussed above, hinder the detection of pathological fear which is not specific to these categories. Moreover, fear conditioning studies employ a large variety of methods to assess fear learning (see Lonsdorf et al., 2017 for an overview). Usually, studies assess subjective behavioral ratings of US probability throughout or after fear learning. These inform about a subject's declarative fear learning and the explicit knowledge about CS-US contingencies. In addition, most studies quantify conditioned responses

with objective, biological measures by applying neuroimaging techniques and /or psychophysiological measures. With either of these measures, it is common to quantify and then contrast responses to CS+ and CS-, with the expectation that CS+ elicits larger fear responses than CS- at the end of fear acquisition.

1.3.1 Fear in neuroimaging research

There are different neurophysiological measures that help inferring on the brain activity associated with fear learning, including electroencephalography (EEG), positron-emission-tomography (PET) or functional magnetic resonance imaging (fMRI). The neuronal circuitry underlying fear conditioning is well investigated in the rodent brain (for a review see Tovote, Fadok, & Luthi, 2015). The amygdala is a bilateral nucleus located in the limbic system, which is suggested to be of central importance for fear learning. This structure can be divided into several sub-nuclei, which maintain different connections to the thalamus, cortical and subcortical brain areas and differentially affect fear expression and inhibition. The basolateral amygdala (BLA) is thought to be the region where information about the CS and the US converges, enabling the formation of an associative fear memory. The dorsal part of the BLA receives inputs from sensory cortices and the thalamus (potentially bearing information about threat cues) and projects to the central nucleus of the amygdala (CEA). The CEA in turn has excitatory projections to the hypothalamus, the periaqueductal grey in the midbrain and the locus coeruleus (LC) in the brainstem, and CEA output can promote autonomic fear responses (for a recent review on the cellular amygdala circuity in associative fear learning see Krabbe, Grundemann, & Luthi, 2017; Tovote et al., 2015).

In the rodent brain, two higher cortical regions are considered of high importance for modulating amygdala output through descending projections: the prelimbic cortex (PL) targets the basal nucleus of the amygdala and is suggested to enhance the expression of fear. The infralimbic cortex (IL) in turn targets inhibitory areas in the amygdala, such as the intercalated (ITC) neurons and the lateral division of the CEA. Input to these amygdala sub-regions inhibits the CEA output and therefore reduces fear expression. In addition, the ventral hippocampus projects to both PL and IL as well as to the BLA and is suggested to be able to modulate fear responses both in an excitatory as well as in an inhibitory way (reviewed by Milad & Quirk, 2012).

This fear circuitry is highly translatable from the rodent to the human brain: while the PL has been suggested to be a homologous region to the human dorsal anterior cingulate cortex (dACC) in fear learning (Milad et al., 2007), the IL is proposed to relate to the human ventromedial prefrontal cortex (vmPFC; reviewed by Milad & Quirk, 2012). Accordingly, dACC and vmPFC appear to have opposing roles in fear enhancement and fear suppression, which is evident in imaging studies of human fear conditioning. Such studies typically contrast brain activation in response to CS+ versus CS-, which reflects brain activity patterns associated with threat or safety signaling. The CS+ > CS- contrast is reliably associated with regions of the salience network (in the context of fear learning also referred to as ‘fear network’), which includes the dACC and bilateral insula, among others (Seeley et al., 2007). The reverse contrast of CS+ < CS- should refer to safety signaling

or fear suppression and typically yields clusters in the vmPFC and the hippocampus (for a meta-analysis see Fullana et al., 2015; for reviews see Greco & Liberzon, 2016; Mechias, Etkin, & Kalisch, 2010; Sehlmeyer et al., 2009).

Even though the central role of the amygdala in fear learning has been studied extensively in animal research, amygdala activity is reported inconsistently in human fear conditioning research. In a review by Sehlmeyer et al. (2009), only 25 out of 44 human fear conditioning studies employing imaging have reported significant amygdala activity in association with the CS+ > CS- contrast. Accordingly, in the recent meta-analysis across 27 imaging studies, amygdala activity did not reach significance for the overall CS+> CS- contrast (Fullana et al., 2015), which is hard to reconcile with its proposed central role in fear learning.

1.3.2 Fear in psychophysiological research

Functional imaging studies typically report the average contrast between CS+ and CS- on the group level. However, fear acquisition and extinction both constitute learning tasks. Averaging across multiple trials and subjects may conceal decisive information about individual differences in the course of fear and safety learning (Lonsdorf & Merz, 2017). In contrast to fMRI, psychophysiological measures allow tracking of the fear learning process more closely by quantifying the magnitude of single physiological responses to CS+ and CS- on the trial level. Instead of targeting brain activity, psychophysiological measures are used to infer indirectly on cognitive processes by capturing changes in the state of the autonomous nervous system. In fear conditioning, these measures are based on the principle that salient or threatening stimuli (such as threat cues) elicit an activating response which gets distributed throughout the body by the autonomous nervous system and is finally detectable in the peripheral physiology.

The brain structures mainly responsible for distributing such activating responses into the periphery are suggested to be the hypothalamus and the LC (Habib, Gold, & Chrousos, 2001; Stratakis & Chrousos, 1995). The hypothalamus regulates the activity of the hypothalamic-pituitary-adrenal (HPA) axis, while the LC is located in the brainstem and its noradrenergic signaling has widespread activating effects throughout the whole brain (Aston-Jones & Cohen, 2005; Samuels & Szabadi, 2008a) as well as on peripheral autonomous arousal (Chrousos & Gold, 1992). Activity in either of these structures enhances sympathetic outflow, while LC activity furthermore inhibits the parasympathetic branch of the autonomous nervous system (Chrousos & Gold, 1992; Samuels & Szabadi, 2008a, 2008b). The resulting output affects several organs such as the heart, the skin, the skeletal muscles or the pupil, enabling the body to respond adaptively to stressors and to orient attention towards motivationally significant stimuli (Habib et al., 2001). Both the LC and the hypothalamus are reciprocally connected with the amygdala (Samuels & Szabadi, 2008a; Tovote et al., 2015), which might modulate their output in particular during fear learning. Hence, salient stimuli or threat cues do not only elicit

specific brain activity patterns, but also cause a physiological response which can be observed in peripheral organs.

Discrete physiological responses to conditioned stimuli can be detected with a variety of psychophysiological methods (for an overview see Lonsdorf et al., 2017). One widely employed measure is the quantification of skin conductance responses (SCR). An increase in autonomous nervous system activity, primarily of the sympathetic branch (Wallin, 1981), is associated with a phasic increase in activity of eccrine sweat glands on the skin of palms and feet (Boucsein et al., 2012). This leads to higher salt concentrations and hence higher conductivity between electrodes placed on the palmar surface. Such phasic increases in skin conductance as a result of stimulation can be quantified (for guidelines see Boucsein et al., 2012). A large body of evidence shows that SCR can be used as a physiological measure of the conditioned response. In human fear conditioning, SCR to CS+ are commonly more frequent and larger in magnitude than SCR to CS- (e.g. Fredrikson & Ohman, 1979; Lipp, Siddle, & Dall, 1998; Luck & Lipp, 2016; see Pineles, Orr, & Orr, 2009 for a list of studies asessing SCR in fear conditioning).

Alternatively to assessing discrete responses to CS+ and CS-, it is also possible to probe physiological responsiveness at different stages during fear learning. Such probes are typically applied during the stimulus interval (i.e. during the presence of CS+ or CS-), as well as during neutral inter-trial intervals. The response magnitude to such probes can then be used to infer on ongoing cognitive-affective processes at the time of probing. Fear conditioning studies often employ sudden loud noises as an acoustic probe, which elicit a defensive startle reflex. This reflex can be quantified with electromyography (EMG) recorded under the eye on the orbicularis oculi muscle (for guidelines see Blumenthal et al., 2005). Startle responses are reliably enhanced during the anticipation of aversive events (as for example during CS+ presentations in anticipation of a US). For this reason, startle EMG is a widely employed physiological measure in fear conditioning studies (e.g. Hamm, Greenwald, Bradley, Cuthbert, & Lang, 1991; Hamm, Greenwald, Bradley, & Lang, 1993; Lipp et al., 1998; Weike, Schupp, & Hamm, 2007) and has also proven useful for extinction and return of fear paradigms (Norrholm et al., 2011; Norrholm et al., 2006).

1.4 Pupillometry as a measure of the conditioned response

1.4.1 Pupillometry as a young measure of conditioned fear

In recent years, pupillometric measures of the conditioned response have gained increasing interest. The diameter of the pupil is determined by the dilator and constrictor muscles of the iris, which are sensitive to sympathetic and parasympathetic input, respectively (Larsen & Waters, 2018; Samuels & Szabadi, 2008a, 2008b). Pupil constriction is elicited by the parasympathetic pathway: the Erdinger-Westphal nucleus in the brainstem innervates peripheral neurons in the ganglion ciliare, which in turn induce pupil constriction

mediated by cholinergic signaling. Input from the sympathetic pathway (arriving via the interomedial cell column of the spinal cord to the superior cervical ganglion) activates the pupillary dilator muscle via noradrenergic signaling. Pupil dilation can be caused by both increasing sympathetic activity as well as by an inhibition of the parasympathetic pathway (and thereby inhibition of constrictor muscle activity). As the LC contributes to sympathetic activity and inhibits parasympathetic outflow (with respective excitatory and inhibitory projections to preganglionic neurons, see Samuels & Szabadi, 2008a, 2008b), LC activity has been found to be closely associated with pupil dilation in the monkey (Joshi, Li, Kalwani, & Gold, 2016), as well as in humans (Alnaes et al., 2014; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Sterpenich et al., 2006). Similarly, hypothalamic activity can affect the pupil size via its sympathetic output: hypothalamus stimulation has been shown to lead to pupil dilation and to inhibit the parasympathetic pupillary light reflex (Loewenfeld, 1958; Sillito & Zbrozyna, 1970).

Due to its sensitivity to autonomous arousal, the pupil responds with dilation to salient or threatening stimuli. For this reason, pupil dilations have gained interest as a measure of the conditioned response, with the first publication appearing 16 years ago (Reinhard & Lachnit, 2002). Most fear conditioning studies employing pupillometry determine the increase in pupil diameter throughout the stimulus interval (Hopkins, Schultz, Hannula, & Helmstetter, 2015; Reinhard & Lachnit, 2002; Reinhard, Lachnit, & Konig, 2006; Visser et al., 2016; Visser, Kunze, Westhoff, Scholte, & Kindt, 2015; Visser, Scholte, Beemsterboer, & Kindt, 2013), while others determine the average pupil diameter during stimulus presentations (De Voogd, Fernandez, & Hermans, 2016). Quantified dilations can then be contrasted between CS+ and CS-. Pupil dilations are usually larger in response to CS+ than to CS- (but see Morriss, Christakou, & van Reekum, 2015) and can hence serve as readout of conditioned fear. Pupil dilations have previously been reported to be maximal in direct US anticipation (Reinhard & Lachnit, 2002). The best CS+ > CS- discrimination may therefore be found when assessing the entire increase in pupil diameter from CS onset to offset.
Another approach is to probe pupillary reflexes during CS presentations, comparable to the application of startle probes. Several studies have shown that the pupillary constriction in response to light probes is less pronounced during states of threat (Bitsios, Szabadi, & Bradshaw, 2002, 2004; Hourdaki et al., 2005).

1.4.2 Pupillometry in comparison to other readouts of conditioned fear

Various psychophysiological measures are thus available to approximate fear learning. When applying such physiological measures with the goal to characterize individual differences in fear processing, it is essential to understand which cognitive-affective and neural processes contribute to different physiological responses. Even though various physiological measures yield the typical CS+ > CS- stimulus difference, they might reflect different aspects of fear learning.

The cognitive processes as well as the neural correlates behind the fear-potentiated of the startle reflex and SCR in fear conditioning have been hypothesized to largely dissociate (Hamm & Vaitl, 1996; Hamm & Weike, 2005; Soeter & Kindt, 2010). SCR are considered as valence-unspecific measure of emotional arousal (Hamm et al., 1993; Lang, Bradley, & Cuthbert, 1990; Lipp, Sheridan, & Siddle, 1994). Contrarily, it has been shown that the startle reflex is modulated by emotional valence, i.e. the reflex is enhanced during the presence of negative, and inhibited during the presence of (equally arousing) positive picture stimuli (Bradley, Cuthbert, & Lang, 1996; Cuthbert, Bradley, & Lang, 1996; Lang et al., 1990; Vrana, Spence, & Lang, 1988). Furthermore, it has been proposed that awareness of CS-US contingencies is crucial for stimulus discrimination measured by SCR (Sevenster, Beckers, & Kindt, 2014; Tabbert et al., 2011; Tabbert, Stark, Kirsch, & Vaitl, 2006; Weike et al., 2007). SCR have previously been related to several higher-order cognitive processes, such as outcome anticipation or conflict monitoring (e.g. Critchley, Mathias, & Dolan, 2001; Kobayashi, Yoshino, Takahashi, & Nomura, 2007; Zeng et al., 2015), which might contribute to SCR magnitude during fear learning. By contrast, the fear-potentiation of the startle response has been proposed to occur even without conscious contingency awareness (Hamm & Vaitl, 1996; Sevenster et al., 2014).

Accordingly, SCR and the startle responses as readouts of fear conditioning are proposed to rely on different neuronal underpinnings. The fear-potentiation of the startle response likely results from a modulation of the reflexive startle circuit via inputs from the amygdala (Davis, 2006; Hamm & Weike, 2005). SCR instead have been suggested to reflect declarative learning and hippocampal activity (Hamm & Weike, 2005) and they have been related to dACC activity and thickness in fear conditioning (Milad et al., 2007).

As pupillometry is a relatively new measure of conditioned fear, it is of interest to relate it to long-established measures such as the fear-potentiated startle reflex or SCR, in order to explore communalities and differences among these physiological measures. Whereas startle responses and SCR commonly habituate over time (Bacigalupo & Luck, 2017; Bradley, Lang, & Cuthbert, 1993; Grillon & Baas, 2003; Pineles et al., 2009), this is unclear for pupillary readouts. Pupil responses to repeated sounds have previously been reported to habituate (Marois, Labonte, Parent, & Vachon, 2017), whereas Hopkins et al. (2015) reported an overall increase of pupil responses during fear conditioning (while simultaneously recorded SCR decreased). Based on such indications for different temporal response patterns, pupil responses may be affected by other cognitive-affective processes than SCR or startle responses during fear learning.

Diverse cognitive processes with the potential to influence the pupil diameter are diverse and they have been reviewed several times (Beatty & Lucero-Wagoner, 2000; Larsen & Waters, 2018; Sirois & Brisson, 2014; van der Wel & van Steenbergen, 2018). The pupil responds for example to emotionally arousing material (Bradley, Miccoli, Escrig, & Lang, 2008; Partala & Surakka, 2003; Snowden et al., 2016), but it is also sensitive to higher order cognitive processes such as prediction and error monitoring processes (Jepma & Nieuwenhuis, 2011; Koenig, Uengoer, & Lachnit, 2017; Preuschoff, Hart, & Einhauser, 2011), uncertainty

(Lavin, San Martin, & Rosales Jubal, 2014) or cognitive effort (e.g. Alnaes et al., 2014; Wendt, Dau, & Hjortkjær, 2016). Such cognitive processes may contribute to pupil dilations during fear learning.

The first part of this work aims at a direct comparison of pupillometric measures of conditioned fear to SCR and startle responses. This comparison is implemented via simultaneous recordings of the three measures during fear learning. We expect all three measures to yield larger responses to CS+ than to CS- after fear acquisition, as this has been demonstrated in previous fear conditioning studies. Apart from this communality, the three measures might display different temporal response patterns throughout the learning process. While we expect SCR and startle responses to habituate, we hypothesize that this is not the case for pupil responses. Given that SCR and startle response have previously been related to different cognitive-affective processes during fear acquisition, we expect to find only a partial overlap of their response patterns. Pupil responses in turn may differ from both other measures. We hypothesize them to reflect rather conscious processes like threat appraisal or US expectancy, for being sensitive to several higher-order cognitive processes and for being maximal in temporal proximity to the US. Pupillary measures throughout fear learning will be characterized in-depth and will be related to SCR and startle responses on a trial-level.

1.4.3 The neural correlates of pupil dilations

Furthermore, the neuronal correlates of pupil dilations during fear learning are unknown. Corresponding to various cognitive processes, different brain activation patterns have been associated with pupillary readouts in concurrent pupillometry and fMRI. Depending on the task under investigation, phasic pupil dilations have been associated with anterior cingulate cortex (ACC) and salience network activity during rest (Murphy et al., 2014; Schneider et al., 2016), with superior temporal gyrus during a speech comprehension task (Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014) or with the dorsal attention network during an attention task (Alnaes et al., 2014). Accordingly, different brain regions may contribute to pupil dilations during fear acquisition and extinction.

As the pupil is responsive to different neuromodulatory circuits, (e.g. noradrenergic or cholinergic signaling, see Larsen & Waters, 2018), it is difficult to directly infer on the cortical processes driving pupil dilations. The LC is a structure that could potentially relay input from higher-order brain regions to the autonomous nervous system, which would then lead to changes in pupil size. The LC receives direct inputs from the anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC) in the monkey brain (Aston-Jones & Cohen, 2005), it projects to the thalamus and it maintains a two-way excitatory connection to the amygdala (Samuels & Szabadi, 2008a). Due to such anatomical considerations, regions of the ‘fear’ or salience network like the dACC, the vmPFC, thalamus or amygdala may therefore influence LC activity and indirectly affect pupil diameter during fear learning. Simultaneous recordings of brain activity (for example measured with fMRI) and recordings of pupil dilations can inform about this relationship during fear learning.

The second part of this work therefore explores the neural correlates of pupil dilations during fear learning with the help of simultaneous assessments of pupillometry and fMRI. Brain activity patterns associated with conditioned pupil responses will be identified and the neural processes associated with changes in tonic pupil diameter are further explored. We hypothesize to find the CS+ > CS- difference represented in pupil response magnitude, as well as in activation of the fear or salience network, as these are commonly reported. We furthermore expect the reverse contrast (CS+ < CS-) to yield activity in vmPFC. If pupil dilations are enhanced in response to CS+, they should correlate positively with salience network activity and negatively with vmPFC activity.

Furthermore, exploring the neural correlates of trial-wise pupil responses within stimulus categories (i.e. after correcting for the general CS+ > CS- stimulus contrast) should reveal brain regions that have a more general, maybe modulatory association with pupil dilations. Candidate regions for this are for example the dACC, the thalamus or the LC, given their central role in the salience network and autonomous response regulation. Furthermore, given that amygdala and regions like dACC and vmPFC are believed to mediate fear learning, these regions could modulate the magnitude of pupil responses via their common associations to the LC.

2. Measuring the conditioned response: a comparison of pupillometry, skin conductance and startle electromyography.

2.1 Summary

In the first study, pupillometry, skin conductance and startle EMG were acquired simultaneously during fear learning (in a two-day fear acquisition, extinction and recall paradigm), in order to relate pupillometry to frequently employed methods in human fear conditioning. As expected, all psychophysiological measures discriminated between CS+ and CS- at the end of the fear acquisition phase, with larger responses to CS+ than to CS-, i.e. each of the measures showed sensitivity to fear learning. We also detected a fear-potentiation of the auditory pupillary reflex to the acoustic startle probes, which constitutes a promising new readout of the conditioned response.

Apart from reflecting the CS+ > CS- stimulus difference, our study results suggest that slow pupil dilations as well as the auditory pupil reflex differ largely from SCR and startle responses during fear acquisition (see Figure 2 B-E). Conditioned responses derived from the different measures correlated only weakly among each other, despite stemming from the same fear learning process. This could be related to significant differences in the temporal response patterns between SCR, startle responses and pupillometric measures during the task. Across trials, pupillary measures increased in response to CS+, whereas SCR and startle responses habituated strongly in response to either stimulus throughout all task phases. Based on these diverging response patterns, we propose that pupillometric readouts reflect partly different cognitive-affective processes than other psychophysiological readouts of the conditioned response.

2.2 Declaration of author contributions

The study was conceived and designed by Victor Spoormaker (VS) and Laura Leuchs (LL). The team of the BeCOME (Biological Classification of Mental Disorders) study at the Max Planck Institute of Psychiatry recruited the participants. LL implemented and programmed the experiment. LL, VS, Max Schneider (MS), Miriam Kraft, Florian Binder and Antonia Motes acquired the data. VS, LL and MS contributed to the interpretation of the results. LL performed the data analysis and took the lead in writing the manuscript with support from VS. All authors provided critical feedback to the manuscript. Dorothee Pöhlchen and Taechawidd Nantawisarakul helped with proofreading of the final manuscript version.

Measuring the conditioned response: a comparison of pupillometry, skin conductance and startle electromyography.

Laura Leuchs, Max Schneider, Victor I. Spoormaker
Max Planck Institute of Psychiatry, 80804 Munich, Germany

Corresponding author: Dr. Victor Spoormaker, Max Planck Institute of Psychiatry, Department of Translational Research in Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany. Telephone: +49 (0) 89-30622-392 Email: spoormaker@psych.mpg.de.

Abstract

In human fear conditioning studies, different physiological readouts can be used to track conditioned responding during fear learning. Commonly employed readouts like skin conductance responses (SCR) or startle responses have in the recent years been complemented by pupillary readouts, but to date it is unknown how pupillary readouts relate to other measures of the conditioned response.

To examine differences and communalities among pupil responses, SCR and startle responses, we simultaneously recorded pupil diameter, skin conductance and startle electromyography (EMG) in 47 healthy subjects during fear acquisition, extinction and a recall test on two consecutive days.

The different measures correlated only weakly, displaying most prominent differences in their response patterns during fear acquisition. Whereas SCR and startle responses habituated, pupillary measures did not. Instead, they increased in response to fear conditioned stimuli and most closely followed ratings of US expectancy. Moreover, we observed that startle-induced pupil responses showed stimulus discrimination during fear acquisition, suggesting a fear-potentiation of the auditory pupil reflex.

We conclude that different physiological outcome measures of the conditioned response inform about different cognitive-affective processes during fear learning, with pupil responses being least affected by physiological habituation and most closely following US expectancy.

1. Introduction

In the face of threat, the autonomous nervous system responds with increased physiological arousal. Such responses are commonly used to track the learning process in human fear conditioning and extinction paradigms. Fear conditioned stimuli (CS+), which predict the occurrence of an aversive unconditioned stimulus (US), typically elicit stronger physiological responses than cues indicating safety (CS-). During extinction, the aversive US is omitted and responses to CS+ typically decline with safety learning. Such conditioned responses can be measured with a variety of methods (i.e. with subjective, neurophysiological and psychophysiological measures; see Lonsdorf et al., 2017 for an overview). Peripheral physiological measures are relatively easy to assess and offer the possibility to track conditioned responses objectively on a single trial level.

One of the most commonly employed psychophysiological measures in human fear conditioning is the quantification of skin conductance responses (SCR). Skin conductance increases in response to emotionally arousing stimuli, proposedly independent of their emotional valence (e.g. Bradley, Miccoli, Escrig, & Lang, 2008; Hamm & Stark, 1993). SCR are typically initiated 1 to 4 s after stimulation and peak between 0.5 and 5 s after initiation (Society for Psychophysiological Research, 2012). In fear conditioning studies, SCR usually yield larger responses to CS+ than to CS- and are often reported to depend on CS-US contingency awareness (Critchley, Mathias, & Dolan, 2002; Sevenster, Beckers, & Kindt, 2014; Tabbert et al., 2010).

Another commonly employed experimental method is to trigger reflexive responses during CS presentations. This way, a startle reflex can be elicited, for example by short bursts of white noise, and its eyeblink component can be recorded with electromyography (EMG) on the orbicularis occuli muscle under the eyes. The muscle contraction in response to startle probes occurs quickly and can be assessed within 120 ms after stimulation (for guidelines see Blumenthal et al., 2005). The magnitude of the startle reflex is reportedly modulated by emotional valence: it is enhanced during negative emotional states and reduced during positive emotional states (Bradley, Cuthbert, & Lang, 1990; for a review see Grillon & Baas, 2003; Lang, Bradley, & Cuthbert, 1990). In fear conditioning studies, the startle response is increased when it is probed during the presence of CS+ as opposed to CS- (Grillon & Ameli, 1998; Hamm, Greenwald, Bradley, Cuthbert, & Lang, 1991). The startle reflex can technically be triggered at any moment, however startle probes are usually applied after several seconds of CS presentation and show strongest fear potentiation in close temporal proximity to the US (Grillon, Ameli, Merikangas, Woods, & Davis, 1993). It is furthermore common to elicit startle responses during neutral inter-trial intervals, to quantify the startle potentiation due to the presence of a CS.

In the last decade, the quantification of pupil dilations as measure of the conditioned response has received increasing interest (e.g. Koenig, Uengoer, & Lachnit, 2017b; Reinhard & Lachnit, 2002; Reinhard, Lachnit, & Konig, 2006; Visser, Kunze, Westhoff, Scholte, & Kindt, 2015; Visser, Scholte, Beemsterboer, & Kindt,

2013). The pupil response can start emerging as early as 200 ms after stimulus onset (Beatty & Lucero-Wagoner, 2000; Sirois & Brisson, 2014) and is modulated by arousal: the pupil has been shown to dilate more strongly to emotionally arousing picture stimuli (Bradley et al., 2008; Snowden et al., 2016) as well as to arousing sounds (Partala & Surakka, 2003) in comparison to neutral stimuli. In fear conditioning studies, the most pronounced CS+ to CS- discrimination has been found in close proximity to the US at stimulus offset (Koenig, Uengoer, & Lachnit, 2017a; Leuchs, Schneider, Czisch, & Spoormaker, 2016; Reinhard & Lachnit, 2002). The conditioned pupil response is often defined as the increase in diameter from stimulus onset to offset, spanning several seconds (e.g. Leuchs et al., 2016; Reinhard & Lachnit, 2002; Visser et al., 2013).

It is yet unknown how pupil responses relate to other physiological readouts of conditioned fear like SCR and startle responses. Despite commonly discriminating between CS+ and CS-, these physiological measures likely capture partly overlapping and partly differing processes throughout fear learning. For instance, it has been hypothesized that startle responses and SCR rely on diverging neuronal circuitry and cognitive processes (for a review see Hamm & Weike, 2005). Startle responses are proposed to relate to reflexive fear circuitry involving the amygdala, while SCR are proposed to rely more on hippocampal activity (Hamm & Weike, 2005) and higher cognitive processes like contingency awareness during fear conditioning (Sevenster et al., 2014). Besides responding to arousal, the pupil has been shown to be sensitive to a variety of higher-order processes like error monitoring (Koenig et al., 2017b; Murphy, van Moort, & Nieuwenhuis, 2016; Preuschoff, Hart, & Einhauser, 2011), uncertainty (Lavin, San Martin, & Rosales Jubal, 2014) or cognitive load (for reviews see Beatty & Lucero-Wagoner, 2000; Sirois & Brisson, 2014; van der Wel & van Steenbergen, 2018). During fear learning it might therefore reflect cognitive processes such as threat appraisal and explicit US expectancy, in addition to arousal.

Moreover, whereas startle responses and SCR are commonly subject to habituation (Bacigalupo & Luck, 2017; Bradley, Lang, & Cuthbert, 1993; Grillon & Baas, 2003; Pineles, Orr, & Orr, 2009), this is unclear for pupillary readouts. We found no substantial habituation of conditioned pupil responses during fear acquisition in a previous study in the MR-environment (Leuchs et al., 2016), whereas other studies have reported habituation of pupil responses to repeated sounds (Marois, Labonte, Parent, & Vachon, 2017). In our previous study, we instead observed robust habituation of the baseline pupil diameter (e.g., the pre-stimulus baseline throughout fear learning), which was inversely correlated to superimposed phasic pupil dilations – in line with previous reports (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Jepma & Nieuwenhuis, 2011). Tonic changes in pupil diameter may therefore reflect habituation, but may also impact the magnitude of pupil responses to stimuli and thus appear to be relevant. To date, some studies have adjusted phasic responses for tonic changes in pupil diameter (e.g. by division or multiplication, see De Voogd, Fernández, &

Hermans, 2016; Marois et al., 2017), while others have subtracted the preceding diameter as a baseline (Reinhard et al., 2006; Visser et al., 2015).

Physiological measures are therefore likely modulated by different cognitive-affective processes, which may be reflected in different temporal response patterns during fear learning. The study aim was to directly compare the properties and temporal dynamics of widely used measures of the conditioned response like SCR and startle responses to pupillary readouts. We simultaneously assessed SCR, startle responses and pupil dilations in healthy subjects during fear acquisition, extinction and a recall test on two consecutive test days. This allowed us to compare different readouts of the conditioned response to the very same CS presentations and in particular to characterize pupillary measures among other, more commonly employed readouts like SCR and startle responses. For each measure, habituation and CS+ to CS- discrimination across task phases were assessed. Furthermore, correlations and differences in the temporal dynamics of conditioned responses were determined across measures. We hypothesized pupil dilations to be less subject to habituation than SCR or startle responses.

2. Method

2.1 Participants

Subjects were recruited at the Max Planck Institute of Psychiatry in Munich in the scope of a large-scale clinical study, which was approved by a local ethics committee. Prior to participation in the study, subjects completed an online questionnaire based on the Composite International Diagnostic Interview Screening Scales (see Kessler et al., 2013) as a short screening. Furthermore, the intake of psychoactive medication or drugs and any former or current psychiatric, neurological or physical diseases were excluded in an interview with a clinician. Fifty-one healthy subjects completed both test days. Four subjects had to be excluded after artifact correction (see below), leaving 47 subjects for the final data analyses (Ø age = 29.5, *SD* = 9.2, 23 female). All subjects had normal or corrected-to-normal vision, gave written informed consent and were reimbursed for participation.

2.2 Materials

2.2.1 Conditioned stimuli

The conditioned stimuli consisted of a square, a circle and a rhombus in three different colors of equal brightness and surface area (for a specification of light conditions and stimulus color-codes see Supplementary Information). One of the conditioned stimuli served as CS- and two served as CS+, while assignment of the three stimuli as CS- or CS+ was counterbalanced across subjects. Stimuli were presented on a black background in the center of a screen, lasted for 4.42 s (CS+) or 4.56 s (CS-; minor differences in

timing were due to software communication delays) and were separated by the presentation of a white fixation cross during the inter-trial interval (duration ranging from 10 to 16 s).

2.2.2 Unconditioned stimuli

The two CS+ were reinforced with two different US: the first US consisted of a 20 ms electrical stimulation implemented with a Linear Isolated Stimulator (Stimsola, BIOPAC Systems, Inc., Goleta, USA). Electrical stimulation was delivered to the right wrist via two Ag/AgCl electrodes filled with electrolyte gel. Shock strength was calibrated individually for each subject by increasing stimulation from zero in 0.5 mA steps, until the subject reported the shock to be very unpleasant and annoying, but not yet painful (Ø = 8.4 mA, *SD* = 4.7, range [2.5 mA; 24.5 mA]). The second US consisted of a 9 bar air blast of 250 ms duration (compare Jovanovic et al., 2005) and was delivered to the larynx from a distance of approximately 1-2 cm.

2.2.3 Startle probe

The startle probe was delivered binaurally via headphones and consisted of a 40ms burst of white noise at 108 decibel with near-instantaneous rise-time (compare Jovanovic et al., 2005). Startle probes were administered at 3.42 s during CS+ trials and at 3.56 s during CS- trials and occurred exactly 1s prior to stimulus offset. Furthermore, startle probes were applied after two thirds of long inter-trial-intervals of 16 s (i.e. at 10.66 s after the previous stimulus offset and 5.33 s before the next stimulus onset).

2.3 Experimental procedures

Subjects underwent uninstructed fear acquisition which was followed by immediate extinction and returned on the next day for a recall task. Before the experiment, subjects were informed that they would be confronted with three aversive stimuli: loud noises presented via headphones, airblasts to the throat and mild electric stimuli to the right wrist. Upon arrival on the first test day (at 10 AM), electrodes were attached and subjects placed their chin on a head-rest, which was located at a distance of 80 cm in front of the screen. Subjects were informed that three geometric shapes would appear alternatingly on the screen and that each could be followed by an airblast or an electric stimulus. They were instructed to view the screen and to find out how the occurrence of the electric shock and the airblast was associated with each the three geometric shapes. Subjects were told that the loud noises were not systematically associated with the stimuli and should be ignored. Whenever a rating was presented on the screen between trials, subjects were asked to report their US probability estimation: for each geometric shape, subjects could first indicate the probability of the occurrence of an airblast with button presses (ranging from 0 to 100 % in 10 % steps). After 18 s, a second rating referring to the shock would appear on the screen for another 18 s.

2.3.1 Habituation

Prior to fear acquisition, subjects underwent an announced habituation phase during which four startle probes were administered, but no US was applied. Each of the three geometric shapes was then presented twice and the rating procedure could be practiced once. After the habituation phase, the experimenter left the room and closed the door.

2.3.2 Fear acquisition

During fear acquisition, the CS- and both CS+ were presented 12 times each in a pseudo-randomized order, with no more than two consecutive presentations of the same stimulus type. In 75 % of fear acquisition trials, the two CS+ co-terminated with their specific US at stimulus offset: the mild electric shock to the wrist was paired with one geometric shape (CS+) while the airblast to the larynx was paired with another shape (CS+air). During fear acquisition, startle probes occurred in 75 % of trials (i.e. nine per stimulus type) as well as during nine inter-trial intervals (see Supplementary Figure S1 for the stimulus order during fear acquisition).

2.3.3 Extinction

Extinction followed fear acquisition immediately after one of the US probability ratings. The CS- and one CS+ (the CS+ previously followed by shocks) were presented ten times each during this task phase, while no US was applied anymore. Six startle probes were applied per stimulus type and another six during inter-trial intervals. At the end of the first test session, subjects were asked to indicate their discomfort with the shock, the airblast and the startle noise on a scale from zero (not at all unpleasant) to ten (extremely unpleasant).

2.3.4 Recall test

Subjects returned at 9 AM on the following day. At the beginning of the session, all electrodes were attached and subjects were once again informed about the procedure. As on the previous day, subjects were instructed to view the screen and to estimate the occurrence of shocks and airblasts in association with the geometric shapes. During recall, the CS- and both CS+ were presented eight times each, while no US was applied. Six startle probes were applied per stimulus type and another six were applied during inter-trial-intervals.

2.3.5 Ratings of US probability

On both test days, ratings of shock and airblast probability were assessed at the beginning, the end, and at intermittent time points throughout the task. Ratings were presented on the screen every 12 trials during fear acquisition and recall and every ten trials during extinction (resulting in a total of seven ratings on the first day and three ratings on the second day). Blocks framed by ratings each contained a balanced amount of stimulus presentations per condition (4 each during acquisition and recall, 5 each during extinction), startle

probes per condition (3 per stimulus type and 3 during inter-trial intervals) and US applications (3 per CS+ during fear acquisition; see Supplementary Figure S1).

2.4 Data acquisition

Eytracking was performed with an EyeLink 1000 desktop system (SR Research Ltd., Ottawa, Canada), which was placed underneath the screen (distance from camera to eye approximately 60cm, resulting in a visual angle of ~7° for the conditioned stimuli). After a standard nine-point-calibration to determine the gaze position on the screen, pupil diameter (in arbitrary units) and gaze coordinates of the right eye were recorded at a sampling rate of 250 Hz. Light conditions were kept constant across all measurements (with bright ceiling lights in a windowless laboratory, see Supplementary Information).

Both skin conductance and startle electromyography (EMG) were tracked with a wireless system (Bionomadics amplifiers and receivers transmitting to a MP150 monitoring system, BIOPAC Systems, Inc., Goleta, USA) and recorded with AcqKnowledge (BIOPAC Systems, Inc., Goleta, USA) at a sampling rate of 1000Hz. Ag/AgCl electrodes with 5mm diameter were filled with electrolyte gel and used for all measurements. For skin conductance recordings, two electrodes were placed on the palm of the left hand, which subjects rested on the desk throughout the experiment. Skin conductance was measured in microsiemens (µS). For startle EMG, two electrodes were placed on the orbicularis oculi muscle underneath the left eye and one electrode was placed behind the left ear for reference. EMG was recorded in microvolts (mV) and the impedance of EMG electrodes was assured to be below 10 kΩ before the recordings were started.

2.5 Data processing

2.5.1 Pupillometry

Eyetracking data was processed and analyzed in Matlab (version 2015a, MathWorks, Natick, USA). First, missing pupil diameter values (due to eye blinks) were replaced by linear interpolation. The last saccade onset before each blink and the first saccade end after each blink marked the beginning and end of the interpolation windows (markers provided by EyeLink software, SR Research Ltd., Ottawa, Canada). Raw data was then smoothed with a 200 ms sliding window and segmented around trials (from 0.5 s before stimulus onset to stimulus offset).

Trial segments were then automatically inspected for three different types of artifacts. First, segments with over 50 % interpolated data points (caused by eye closure or excessive blinking) were excluded (2.1 % of all trials, *SD* = 4.1 %). Two subjects were excluded due to excessive blinking (resulting in over 30% missing trials on each test day). Second, trials in which the subjects' gaze was not directed at the middle of the screen for over 0.5 s were discarded (5.9 % of all trials, *SD* = 4.7 %). For this purpose, a cutoff window was defined

around the subject's median gaze position (median across all trials). The limits of the cutoff window were informed by the mean gaze deviation during trials across all subjects (approximately 14 % of the visual angle). Third, trials containing sudden shifts in pupil diameter, which are rather due to recording artifacts than biological processes, were identified. For each subject, the standard deviation across all trials was determined for five different time epochs: one epoch of a 0.5 s pre-stimulus baseline and following four epochs of each 1 s duration, covering the stimulus presentation. Trials deviating by more than 3.3 SD from all trials of that subject in any of these epochs were excluded (4.0 % of all trials, *SD* = 1.9 %). Artifacts identified by these three approaches partly overlapped, resulting in a total of 11.4 % excluded trials (*SD* = 6.3 %) and were treated as missing values (corresponding to on average 9.1 out of 80 trials across both test days).

For the quantification of pupil response amplitudes, a window of interest in close US-proximity was chosen. Slow pupil dilations in response to the conditioned stimuli were calculated by subtracting the pupil diameter at pre-stimulus baseline (average across 0.5 s before stimulus onset) from the maximum pupil diameter in the last second before stimulus offset (see Figure 1C). In an additional analysis, we assessed auditory pupil responses to the startle probes. This was done to test whether the auditory pupil reflex discriminated between CS+, CS- and inter-trial intervals. For each startle probe (during stimuli and inter-trial intervals), the change in pupil diameter from startle onset to the maximum dilation within 1 s was quantified. For analyses of overall changes in tonic pupil diameter, pre-stimulus baselines were used. All raw pupil responses were separately z-transformed within subjects (across all trials of both test days).

2.5.2 Skin Conductance

Skin conductance data was low pass filtered at 1 Hz in AcqKnowledge (BIOPAC Systems, Inc., Goleta, USA) and further processed in Matlab (version 2015a, MathWorks, Natick, USA). For each trial, the maximum skin conductance between 2.5 and 4.5s after stimulus onset was identified. To determine the SCR amplitude per trial, the segment preceding the maximum response was searched for a local minimum (between 1s after stimulus onset and the time of the maximum response). If there was an identifiable trough, the response amplitude was defined as the minimum to maximum difference (see Figure 1B), otherwise (i.e., if skin conductance only declined during the trial) the response amplitude was set to zero.

Furthermore, trials were automatically inspected for outliers by examining the standard deviation of each trial segment in comparison to all other segments of the same subject. If the trial's standard deviation exceeded the mean standard deviation across all trials over 3.3 times, the respective trial was treated as missing value (on average 8.8 % of all trials, *SD* = 1.6 %; corresponding to on average 7.0 out of 80 trials across both test days). Two subjects were excluded as non-responders for having over two thirds (67 %) of zero responses on each test day. All SCR to stimulus onsets were z-transformed across both test days within subjects.

2.5.3 Startle EMG

Startle EMG data was band pass filtered between 28 and 400 Hz, then rectified and low pass filtered at 40 Hz (see Blumenthal et al., 2005) in AcqKnowledge (BIOPAC Systems, Inc., Goleta, USA). All further data processing was implemented in Matlab (version 2015a, MathWorks, Natick, USA). Startle responses were calculated by subtracting the average baseline (50 ms before the startle probe) from the maximum amplitude in the window of interest from 20 to 120 ms after the startle probe (see Figure 1A). Trials in which either the standard deviation or the maximum amplitude during the baseline exceeded the standard deviation or maximum during the window of interest were identified as invalid and treated as missing values. This resulted in a total of 5.7 % excluded trials (*SD* = 5.6 %; corresponding to on average 4.7 out of 82 trials across both test days). No subjects were excluded after artifact correction (none with more than 30 % invalid trials). All raw startle amplitudes from both test days were z-transformed per subject.

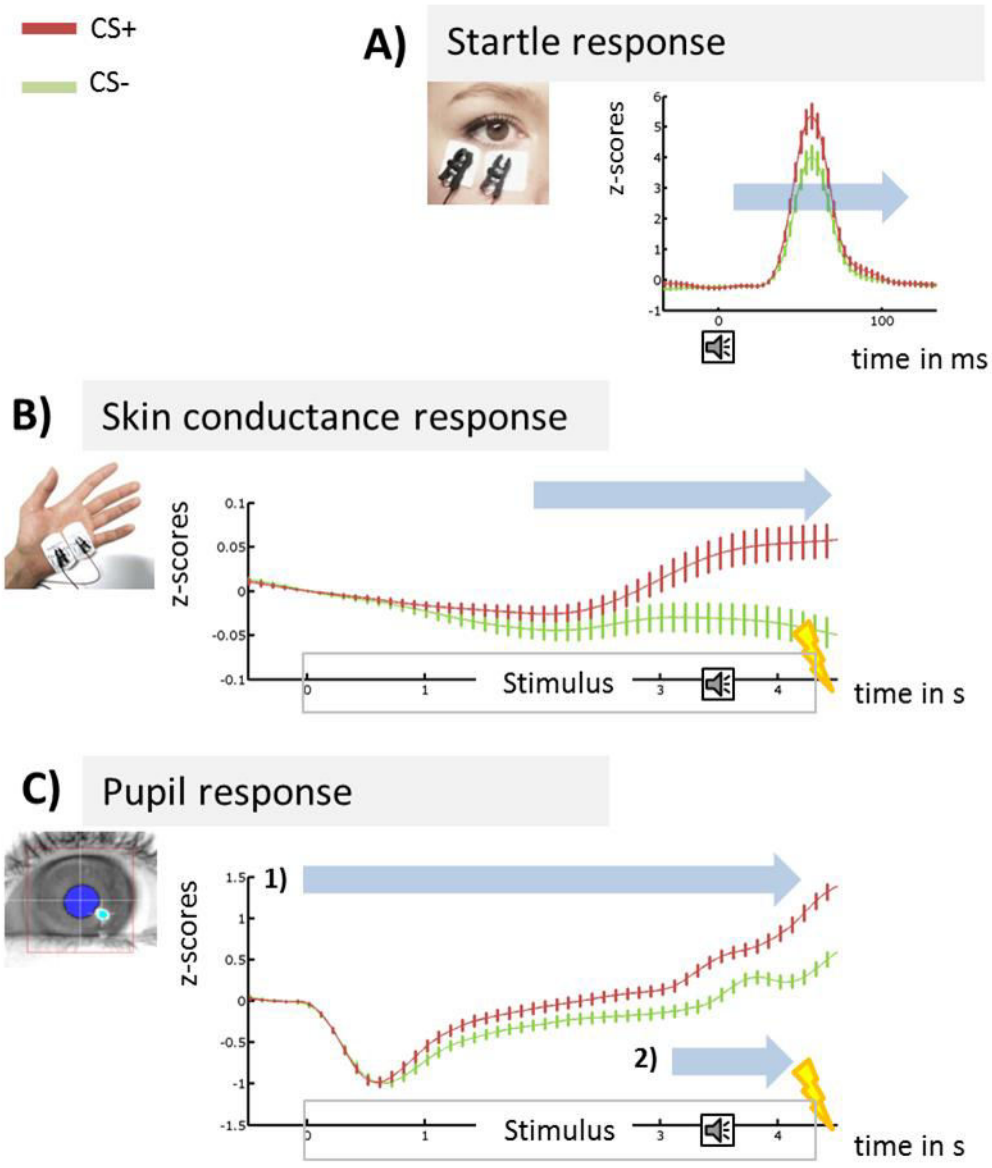


Figure 1. Physiological responses to CS+ (red) and CS- (green), averaged across subjects during the last block of fear acquisition. Vertical lines indicate the standard error of the mean (SEM). Blue arrows indicate the windows of interest for the different physiological readouts: A) Startle response amplitude: difference from baseline (mean 50 ms before startle probe) to peak (maximum during 20 to 120 ms after the startle probe). B) Skin conductance response (SCR) amplitude: difference from local minimum to peak (trough after stimulus onset to maximum during 2.5 to 4.5 s after stimulus onset). C) Pupil responses 1) slow pupil dilations: difference from baseline (mean 0.5 s before stimulus onset) to peak in the last second before stimulus offset 2) auditory pupil reflex to the startle probe: increase in pupil diameter from probe onset to the maximum within 1 s. Depicted segments were derived from z-transformed physiological data. Pre-stimulus baselines were subtracted prior to averaging.

2.6 Statistical Tests

For simplification, and because only this specific CS+ was presented during extinction, all further analyses concerning the CS+ > CS- difference will focus only on the CS+ that was reinforced by electric shocks during fear acquisition (results for the CS+ coupled with airblasts are presented in Supplementary Table S1; comparisons between the different CS+ are presented in Supplementary Table S2). To compare the exact same trials across measures, we furthermore only included trials with startle probes in the following analyses. After artifact correction, 47 out of 51 subjects were included in the final data analyses. Statistical testing was performed with SPSS (PASW Statistics for Windows, Version 18.0. Chicago, SPSS Inc.) and Matlab (version 2015a, MathWorks, Natick, USA).

2.6.1 Ratings of US probability

To assess explicit learning of CS-US contingencies, we performed a repeated measures analysis of variance (rmANOVA) across ratings of shock probability for each test phase with the factors stimulus (CS- and CS+) and time (four ratings until the end of fear acquisition, two ratings during extinction, three ratings during recall).

2.6.2 Physiological responses

Z-transformed response amplitudes were used for testing discrimination between CS+ and CS- and for the exploration of trial-wise dynamics, thereby disregarding inter-individual differences in absolute response magnitude. To assess the temporal dynamics of physiological responses, we performed separate rmANOVAs for each measure and test phase. To overcome the problem of missing values (originating from artifact correction) while yet preserving information about the temporal dynamics, trials between consecutive ratings were averaged in a block-wise manner. For each subject, this procedure resulted in three values for fear acquisition, two values for extinction and two values for the recall phase per stimulus (CS+ and CS- each; every value consisted of an average of three physiological responses from trials including startle probes). These values were entered into one rmANOVA per task phase, each with the factor stimulus (two) and time (three values for acquisition and two values for extinction and recall, respectively). For each measure, the main effects of stimulus, time, as well as stimulus × time interactions were determined. If the assumption of sphericity was not met in a given rmANOVA, a Greenhouse Geisser correction was applied (uncorrected degrees of freedom are reported).

Furthermore, for each measure and each task phase, paired t-tests were performed across averaged responses to CS+ and CS-. In a similar vein, t-tests were performed between averaged responses to startle probes during inter-trial-intervals versus CS- and CS+ presentations (for startle responses and auditory pupil responses). As

an estimate of the overall effect size of the mean CS+ > CS- difference, Cohen's d was calculated for each task phase and each outcome measure (mean CS+ to CS- difference, divided by the average standard deviation of all responses).

2.6.3 Pupil baseline analyses

Pre-stimulus pupil baselines were used to assess changes in tonic pupil diameter over time and to determine their relationship with superimposed phasic dilations. For these analyses, baselines preceding all trials with startle probes (also CS+ air) were included. We calculated one Pearson correlation per subject between pre-stimulus baselines and respective pupil responses (from 57 trials including startle probes: 39 stemming from the first and 18 from the second test day). Resulting r-values were tested for significance with a non-parametric Wilcoxon signed rank test. To determine if there was a significant effect of time, baselines were averaged block-wise between the ratings (irrespective of stimulus type) and entered into one rmANOVA per test day. The rmANOVAs accordingly comprised five values for the first test day (each an average of nine baselines per fear acquisition block and of six baselines per extinction block) and two values for the second test day (each an average of nine baselines per block).

2.6.4 Correlations and comparative analyses between physiological outcome measures

To compare the temporal dynamics of pupillary readouts with the other two measures in particular, we performed two rmANOVAs each containing startle responses, SCR and one pupillometric readout (i.e. slow pupil dilations or auditory pupil responses to startle probes). These rmANOVAs were used to determine interactions of measure × time and measure × stimulus × time and thereby test for statistical differences in the temporal response patterns across measures.

To explore the association between SCR, startle probes, and pupillary readouts on a trial-level, Pearson correlations were calculated for all pairs of the four measures. This was done across all trials, as well as for fear acquisition and recall separately. To maximize the amount of available data points, we did not exclude outlier values for these analyses. To control for the effect of outlier values, we additionally calculated Spearman rank correlations. To test for significance, resulting r-values were tested for significance with Wilcoxon signed rank tests. Furthermore, we computed Pearson correlations between the three measures across averaged CS+ to CS- difference scores during fear acquisition, to test if CS+ to CS- discrimination was correlated between measures.

3. Results

3.1 Ratings of US probability

Ratings of US probability revealed significant stimulus discrimination for all task phases, with significantly higher ratings of US probability for CS+ (see Table 1; see Supplementary Table S1 for the second US). At the end of the fear acquisition phase (fourth rating), the mean shock probability was rated as 81% for CS+ (*SD* = 18%) and as 6% for CS- (*SD* = 19%). Ratings of discomfort revealed that subjects perceived the startle probes to be equally aversive as the electric shocks, and more aversive than airblasts (ratings on a scale from 0 to 10; mean shock = 6.9, *SD* = 1.5, mean startle = 7.0, *SD* = 2.3, mean airblast = 5.2, *SD* = 2.8; see Supplementary Figure S2 for ratings).

3.2 Physiological responses

The results of the rmANOVA, t-tests and effects sizes for all physiological measures are reported in Table 1. During fear acquisition, SCR and startle responses showed a significant stimulus effect with larger responses to CS+ than CS- (all $p < .01$) and stimulus × time interactions (startle responses at trend with $p = .08$). To ensure that we did not capture the SCR to startle probes, which (with a minimum onset latency of 1 s, see Boucsein et al., 2012) could theoretically start at 4.42 s, we additionally analyzed peaks with a more conservative cut-off at 4.0 s. This yielded highly similar results: SCR values derived from both scoring methods correlated very highly (mean r across subjects of .98, *SD* = .02), and the t-values for the CS+ > CS- comparison of all acquisition trials were very similar ($t(46) = 5.71$ versus $t(46) = 5.67$).

SCR and startle responses revealed significant effects of time for all task phases (all $p < .01$, only SCR during extinction did not reach significance). For both measures, post-hoc t-tests revealed a significant decline in response to both CS+ and CS- from the first to the last block of fear acquisition (and also for recall, all $t > 2.9$, all $p < .005$; compare Figure 3). Slow pupil dilations did not yield a significant effect of stimulus or time, but a significant stimulus × time interaction during fear acquisition. Post-hoc t-tests revealed a significant CS+ > CS- difference for the last block of fear acquisition ($t(46) = 3.03$, $p < .005$), and that pupil responses to CS+ stimuli increased at trend level from the first to the last block of fear acquisition ($t(46) = 1.46$, $p = .06$; compare Figure 3). The auditory pupil reflex to startle probes yielded a significant effect of stimulus during fear acquisition, with on average larger responses to startle probes presented during CS+ than during CS- ($t(46) = 2.59$, $p < .01$; see Figure 2). The effect of time and the stimulus × time interaction did not reach significance, but post-hoc t-tests revealed a significant increase in response to CS+ startle probes from the first to the last block of fear acquisition ($t(46) = 2.83$, $p < .005$). During extinction, none of the physiological measures showed significant stimulus discrimination; during recall, only SCR and slow pupil dilations yielded larger responses to CS+ than CS-.

Table 1. Stimulus effect and temporal analysis for all outcome measures: results of paired t-tests (CS+ > CS-), estimates of the stimulus effect (Cohen's *d*) and results of rmANOVAs for ratings of shock probability and block-wise averages of physiological responses; rmANOVAs contained the factors stimulus (CS- & CS+) and time (three values for fear acquisition, two values for extinction and two values for recall).

	t-test CS+ > CS-	Cohen's d	Stimulus	Time	Stimulus × Time
Ratings of shock probability					
Fear acquisition	t (45) = 14.80, p < .001	2.16	F(1,46) = 240.17, p < . 001	F(3,138) = 4.12, p < .05	F(3,138) = 59.09, p < . 001
Extinction	t (45) = 2.27, p < .001	.33	F (1,46) = 5.16, p < .05	F (1,46) = 1.55, p = .22	F(1,46) = 1.55, p = .22
recall	t (45) = 5.24, p < .001	.84	F (1,46) = 32.92 p < .001	F (2,92) = 199.34 p < .001	F (2,92) = 25.55, p < .001
SCR					
Fear acquisition	t (45) = 5.27, p < .001	.75	F (1,42) = 23.21, p < .001	F (2,84) = 20.46, p < .001	F (2,84) = 4.35, p < .05
Extinction	t (45) = 0.81, p = .21	.12	F (1,45) = 1.22, p = .28	F (1,45) = 2.00, p = .16	F (1,45) = 2.38, p = .13
recall	t (45) = 3.39, p < .001	.49	F (1,43) = 7.21, p < .05	F (1,43) = 30.69, p < .001	F (1,43) = 0.00, p = .97
Startle response					
Fear acquisition	t (46) = 2.78, p < .005	.41	F (1,46) = 7.73, p < .01	F (2,92) = 58.01, p < .001	F (2,92) = 2.55, p = .08
Extinction	t (46) = 0.56, p = .17	.08	F (1,46) = 0.32, p = .58	F (1,46) = 8.12, p < .01	F (1,46) = 0.06, p = .80
recall	t (46) = 1.53, p = .07	.22	F (1,46) = 0.95, p = .34	F (1,46) = 63.60, p < .001	F (1,46) = 7.19, p < .05
Slow pupil dilations					
Fear acquisition	t (46) = 0.98, p = .17	.14	F (1,44) = 0.45, p = .51	F (2,88) = 0.01, p = .99	F (2,88) = 3.21, p < .05
Extinction	t (46) = 0.44, p = .33	.06	F (1,43) = 0.09, p = .77	F (1,43) = 3.01, p = .09	F (1,43) = 2.19, p = .15
recall	t (46) = 2.90, p < .005	.42	F (1,44) = 9.24, p < .005	F (1,44) = 10.46, p < .005	F (1,44) = 1.36, p = .25
Auditory pupil response to the startle probe					
Fear acquisition	t (1,46) = 2.59, p < .01	.38	F (1,44) = 6.86, p < .05	F (2,88) = 2.33, p = .10	F (2,88) = 2.15, p = .12
Extinction	t (1,46) = 0.82, p = .21	.12	F (1,43) = 0.37, p = .55	F (1,43) = 7.01, p < .05	F (1,43) = .07, p = .80
recall	t (1,46) = 0.20, p = .42	.03	F (1,44) = 0.02, p = .90	F (1,44) = 7.88, p < .01	F (1,44) = 0.00, p = .98

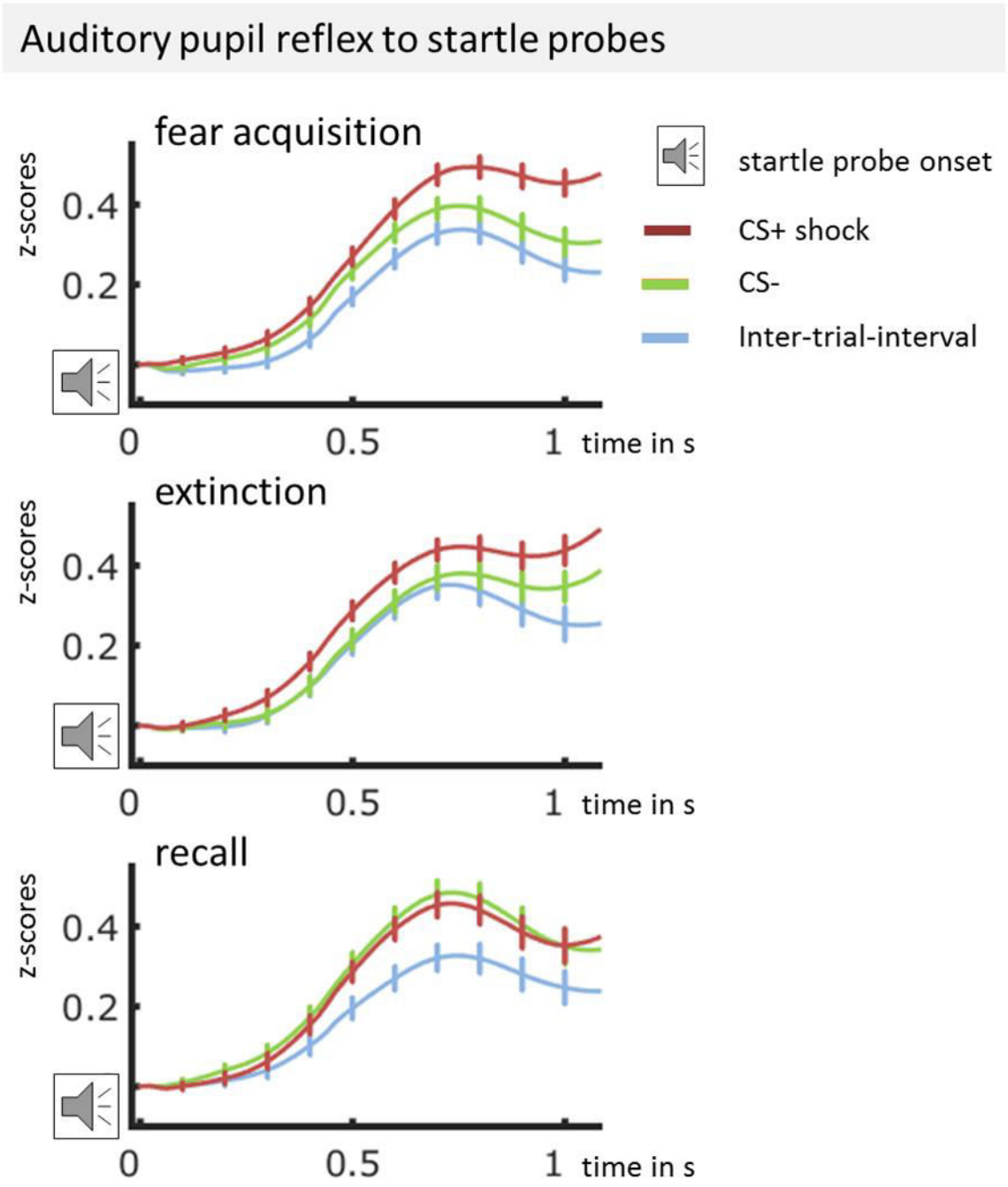


Figure 2. Auditory pupil responses to startle probes applied during inter-trial intervals (blue), CS- (green) and CS+ (red) during fear acquisition, extinction and recall. Probe onset occurred at 0 s. Vertical lines: SEM. Depicted segments were derived from z-transformed physiological data. Pre-stimulus baselines were subtracted prior to averaging.

T-tests revealed significantly larger responses to startle probes applied during stimuli (CS+ as well as CS-) than to probes applied during inter-trial intervals for both startle responses and auditory pupil responses (except auditory pupil responses during extinction, see Table 2 and Figure 2). The screen luminance differed between stimulus presentations and inter-trial-intervals, which could affect the baseline pupil size and, relatedly, the pupil response. Therefore, a post-hoc t-test was performed between the respective average pupil baselines at startle onset (inter-trial intervals vs. stimulus presentations), which was not significant (t (46) = 0.55, p = .59, Cohen's d = .08).

Table 2. Difference in response magnitude to startle probes applied during inter-trial intervals (ITIs) and stimuli: results of paired t-tests CS- > ITI and CS+ > ITI for startle responses and auditory pupil responses to the startle probe for the task phases fear acquisition, extinction and recall.

	t-test CS- > ITI	**t-test CS+ > ITI**
Startle response		
Fear acquisition	t (46) = 3.19, p < .005	t (46) = 5.20, p < .001
Extinction	t (46) = 3.35, p < .005	t (46) = 3.90, p < .001
recall	t (46) = 2.29, p < .05	t (46) = 2.99, p < .005
Auditory pupil response to the startle probe		
Fear acquisition	t (46) = 1.69, p < .05	t (46) = 3.67, p < .001
Extinction	t (46) = 0.58, p = .28	t (46) = 1.38, p = .09
recall	t (46) = 1.63, p = .05	t (46) = 1.99, p < .05

3.3 Pupil baseline analyses

Pre-stimulus pupil baselines yielded a significant effect of time on both test days (first test day: F(4,184) = 37.23, p < .001; second test day: F(1,46) = 24.51, p < .001). Post-hoc t-tests revealed a significant decline in pupil baseline diameter from the first to the last block on each day (first test day: t (46) = 7.74, p < .001, second test day: t (46) = 4.95, p < .001). As in previous studies, we found a moderate negative correlation of phasic pupil response magnitude with the preceding baseline diameter (slow pupil dilations: range of r = [-.83; .24], mean r = -.42; p < .001, see also Supplementary Figure S3; auditory pupil responses: range of r = [-.65; .22], mean r = -.22; p < .001).

3.4 Correlations and comparative analyses between physiological outcome measures

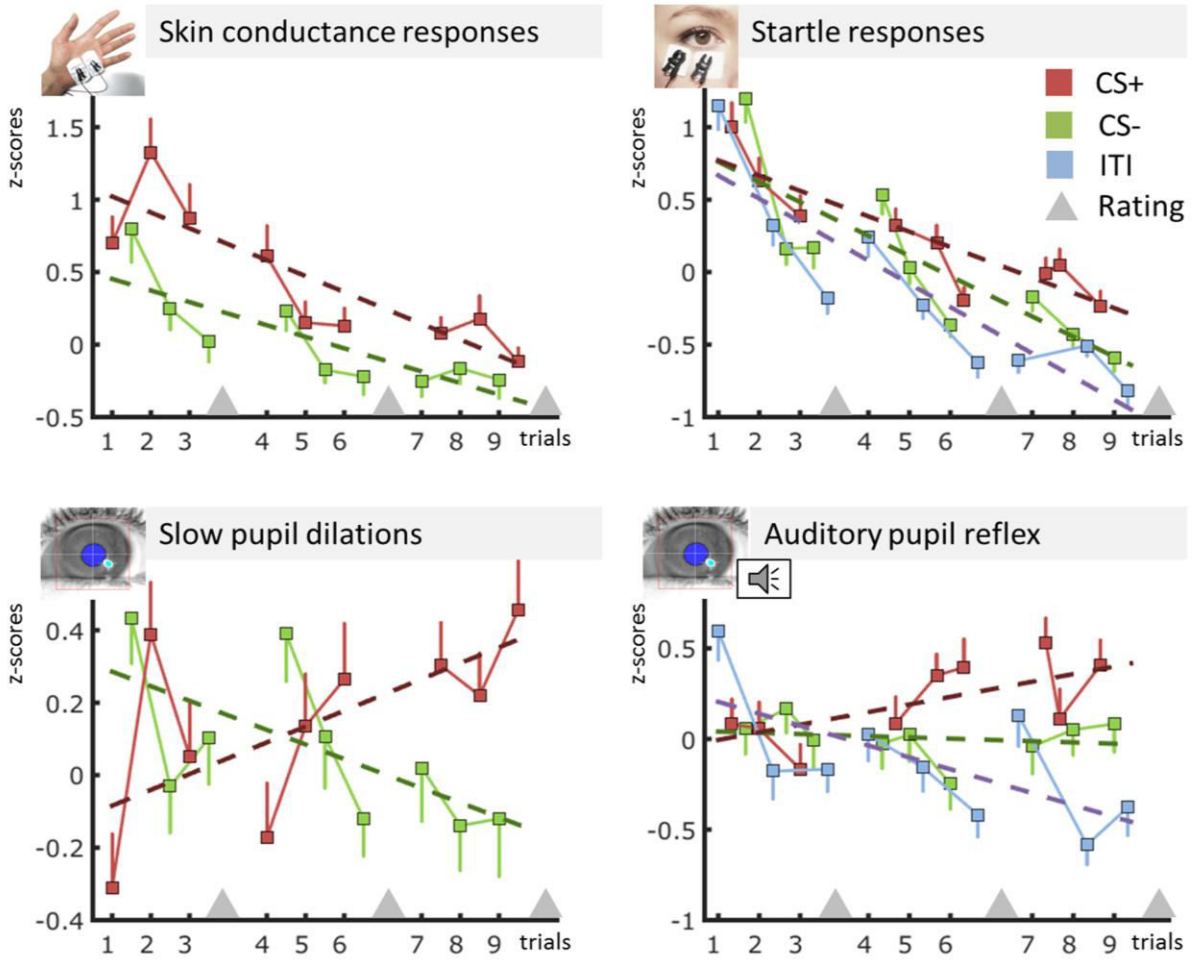


Figure 3. Physiological responses throughout fear acquisition. Responses to CS+ or startle probes applied during CS+ are depicted in red, responses to CS- or startle probes applied during CS- in green, responses to startle probes applied during inter-trial intervals (ITIs) in blue. Dashed lines: linear regression; vertical lines: SEM (one-sided).

The differences in temporal dynamics between pupillary readouts and the other measures were evident in the two rmANOVAs containing SCR, startle responses and one pupillary measure, respectively (i.e. either slow pupil dilations or auditory pupil responses). During fear acquisition, both rmANOVAs yielded a significant measure × time interaction (for slow pupil dilations: $F(4,160) = 8.32$, $p < .001$; for auditory pupil responses: $F(4,152) = 13.36$, $p < .001$), as well as a significant measure × stimulus × time interaction (for slow pupil dilations: $F(4,160) = 3.83$, $p < .01$; for auditory pupil responses: $F(4,152) = 4.29$, $p < .005$; for main effects and interactions for all task phases see Supplementary Tables S3 and S4). To further disentangle the significant measure x time interaction, we computed the decline in response magnitude from the beginning to

the end of fear acquisition for each measure (by subtracting the average across stimuli of the last block from the average across stimuli of the first block of fear acquisition). Post-hoc t-tests revealed that the decline in SCR (mean = -.72, *SD* = .76) and startle responses (mean = -.90, *SD* = .55) did not differ significantly (t (46) = 0.27, $p > .10$). Both SCR and startle responses declined more strongly than either of the pupillometric measures (all t (46) > 4.50 and all $p < .001$; slow dilations: mean = -0.05, *SD* = 0.56; auditory pupil reflex: mean = 0.12, *SD* = 0.52). There was only a trend for a difference in decline between the pupillometric readouts (t (46) = 1.72, p = .08).

Pearson correlations on a trial-level between different measures were weak, but mostly significant (significant correlations with $p < .05$ ranged from r = .08 to r = .27), with highest correlation values for recall trials (see Table 3; Spearman correlations yielded highly similar results). Auditory pupil responses only correlated significantly with (overlapping) slow pupil dilations. Average CS+ > CS- difference scores (calculated for fear acquisition) yielded mixed results: interestingly, startle difference scores correlated higher with slow pupil dilations (r = .41, $p < .005$) than with auditory pupil responses to the startle probe (r = .25, p = .08). Difference scores from both pupillary measures were inter-correlated (r = .30, $p < .05$), whereas SCR difference scores did not correlate significantly with any other measure (with slow pupil dilations at trend: r = .28, p = .06; with the pupil reflex to startle probes: r = .06, p = .68; with startle responses: r = .19, p = .20).

Table 3. Trial-wise correlations across measures. Correlations were determined per person; reported are the mean r-value and the significance level of a Wilcoxon signed rank tests across individual r-values. Correlations between measures are provided for all trials (57 physiological responses across both test days) and separately for fear acquisition and recall trials. *indicates $p < .05$, ** indicates $p < .001$.

	SCR	**Startle response**	**Slow pupil dilations**
Startle response			
All trials	r = .18**		
Fear acquisition	r = .16**		
Recall	r = .21**		
Slow pupil dilations			
All trials	r = .13**	r = .15**	
Fear acquisition	r = .11**	r = .08*	
Recall	r = .12*	r = .27**	
Auditory pupil response to the startle probe			
All trials	r = -.03	r = .04	r = .20**
Fear acquisition	r = -.05	r = .03	r = .18**
Recall	r = -.01	r = .07	r = .20**

4. Discussion

We conducted simultaneous measurements of skin conductance, startle EMG and pupil diameter during a two-day fear acquisition, extinction and recall paradigm. The study aim was to provide a direct comparison of the three measures' properties and the temporal dynamics of conditioned responses throughout the tasks, with an in-depth examination of pupillary readouts. US expectancy ratings indicated that subjects had learned reinforcement rates. Furthermore, all physiological readouts yielded larger responses to CS+ than CS- during fear acquisition, with comparatively late stimulus discrimination for slow pupil dilations. The main difference between physiological readouts was apparent in their temporal dynamics throughout fear acquisition and was reflected in significant interactions of measure by time. Whereas SCR and startle responses showed strong habituation, pupillary readouts instead increased in response to CS+, thereby roughly following ratings of US expectancy throughout fear learning.

4.1 Relating physiological outcome measures to cognitive-affective processes

4.1.1 Pupil responses

Slow pupil dilations discriminated between CS+ and CS- in late fear acquisition and during recall. They have previously been found to approximate reinforcement rates by yielding stronger responses to fully than to partially reinforced CS+ during initial fear learning (Koenig et al., 2017b; Leuchs et al., 2016; see Supplementary Figure S4). Koenig et al. (2017b) found this pattern to reverse in later learning stages, with stronger pupil responses to intermittently reinforced CS+ than fully reinforced CS+. They concluded that pupil dilations reflect continuous updating of US prediction. Uncertainty about stimulus outcome may hence contribute to pupil dilations in studies with partial reinforcement.

In a previous study, we found slow pupil dilations during fear learning to be associated with activity in regions of the salience network (see Fullana et al., 2015 for a meta-analysis of brain activity associated with fear conditioning; Leuchs et al., 2016). Pupil dilations robustly related to dorsal anterior cingulate cortex activity on a trial-level during fear acquisition (Leuchs et al., 2016). However, we also observed similar neural correlates of pupil dilation during a reward anticipation task (Schneider, Leuchs, Czisch, Samann, & Spoormaker, 2018). This indicates that slow pupil dilations in fear learning may largely reflect valence-independent outcome expectancy and its updating, thereby showing similarity with ratings of US probability.

In our study, we used a partial reinforcement schedule (75%) for two different CS+. The relatively high task complexity (with two partially reinforced CS+), as well as the application of startle probes (Sjouwerman, Niehaus, Kuhn, & Lonsdorf, 2016) may have prolonged the explicit learning of CS-US contingencies, and may have increased overall uncertainty. This might explain why slow pupil dilations started discriminating rather late between CS+ and CS- throughout fear acquisition, and may account for the prolonged stimulus discrimination during recall.

We furthermore observed a fear-potentiation of the auditory pupil reflex to startle probes during fear acquisition. To our knowledge, this is the first report of this phenomenon in fear conditioning. Auditory pupil responses have previously been found to be larger for arousing than neutral sound stimuli (Partala & Surakka, 2003; Widmann, Schroger, & Wetzel, 2018). Interestingly, Widmann et al. (2018) found that the initial component of this auditory pupil response (within approximately 1 s after sound onset) was related solely to stimulus novelty and not to the arousing intensity of the presented sound material. However, in our case, the auditory pupil response was not elicited by emotionally arousing material, but was triggered during the anticipation of a US (or safety for CS-). The magnitude of this triggered auditory pupil response may therefore reflect the state of arousal at the time of probing, resulting in discrimination between CS+, CS- and inter-trial interval probes.

Other pupillary reflexes, triggered during US anticipation, have previously served as measures of conditioned fear: it has been demonstrated that the pupil constriction in response to light flashes is inhibited during threat anticipation (Bitsios, Szabadi, & Bradshaw, 1996; Hourdaki et al., 2005). This inhibition of the pupillary light reflex was also found to follow emotional arousal rather than valence (Bitsios, Szabadi, & Bradshaw, 2004; Henderson, Bradley, & Lang, 2014) and to rely on different circuitry than the acoustic startle reflex (by being susceptible to different pharmacological manipulations; see Bitsios, Philpott, Langley, Bradshaw, & Szabadi, 1999).

In our study, auditory pupil responses yielded only small trial-wise correlations with slow pupil dilations (even though the analysis intervals were partially overlapping and therefore not entirely independent). The auditory pupil responses may be less subject to higher cognitive processes (such as uncertainty or monitoring of CS-US contingency) than slow pupil dilations, due to being triggered and of reflexive nature. However, auditory pupil responses also did not display much similarity with other reflexive physiological readouts like startle responses or SCR at CS onset by not correlating significantly with these measures on the trial-level and by displaying no habituation. The auditory pupil response to startle probes may hence serve as a complementary readout of fear learning which differs from other physiological measures.

4.1.2 Pupil baseline measurements

The lack of habituation in pupil responses might be specific for emotionally arousing states: whereas pupil dilations showed no habituation in response to repeatedly presented emotional pictures in one study by Snowden et al. (2016), they have been reported to habituate in response to neutral sounds (Marois et al., 2017; see Nieuwenhuis, De Geus, & Aston-Jones, 2011 for a discussion of pupil dilations as orienting response). We suggest that physiological habituation during fear learning is not reflected in phasic pupil responses, but instead in the tonic decline of the pupil baseline diameter. This decline was clearly present on both test days, as well as in another fear conditioning study in the MR environment (Leuchs et al., 2016). Tonic changes in pupil diameter have previously been proposed to track wakefulness and general alertness

(Gilzenrat et al., 2010; Lowenstein, Feinberg, & Loewenfeld, 1963; Wilhelm, Ludtke, & Wilhelm, 1998). Accordingly, we have previously found that the reduction in baseline-weighted pupil dilations is associated with declining thalamic activity (Leuchs et al., 2016).

In line with previous reports (Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011; Leuchs et al., 2016), the baseline pupil diameter furthermore correlated negatively with the magnitude of subsequent, superimposed responses (r = -.42 for slow pupil dilations and r = -.22 for auditory pupil responses). The causality of this relationship remains to be explored, for instance by clarifying whether simple floor and ceiling effects limit additional constriction or dilation on top of a given pupil diameter. It is possible to correct or weight pupil responses by their preceding baseline diameter, for instance by multiplication (Leuchs et al., 2016) or division (Marois et al., 2017); however, to our knowledge, there is no established standard procedure. We propose that changes in tonic baseline diameter should be reported together with pupil responses and that different experimental conditions should be carefully balanced over time to avoid systematic differences in baseline diameter (and therefore superimposed phasic responses).

4.1.3 SCR and startle responses

Several studies have compared startle responses and SCR in fear learning. As opposed to startle responses, SCR are considered a valence-unspecific measure of emotional arousal (Hamm, Greenwald, Bradley, & Lang, 1993; Lang et al., 1990; Lipp, Sheridan, & Siddle, 1994) and SCR-based CS+ to CS- discrimination proposedly depends on CS-US contingency awareness (Tabbert et al., 2011; Tabbert, Stark, Kirsch, & Vaitl, 2006). By contrast, the fear-potentiation of the startle response has been proposed to depend less on contingency awareness (Hamm & Vaitl, 1996; Sevenster et al., 2014), but to be modulated by emotional valence (Lang et al., 1990). In a review, Hamm and Weike (2005) suggest that the startle reflex and SCR rely on two dissociable processes during fear learning: one subcortical, amygdala-dependent system, which does not depend on explicit contingency awareness and modulates reflexive fear responses (proposedly the startle reflex), and one system which is responsible for declarative learning of CS-US contingencies and is presumably hippocampus-dependent (manifest in SCR).

Such a clear distinction of processes underlying startle responses and SCR may however not apply to all cases. It has been suggested that contingency awareness has an influence on successful fear inhibition as measured by startle response magnitude (Jovanovic et al., 2006). Moreover, results from a study by Purkis and Lipp (2001) indicate that startle potentiation does require awareness of CS-US contingencies when contingency detection is made more difficult (see also Lipp & Purkis, 2005). Furthermore, the valence-specificity of the startle response has primarily been demonstrated for emotional picture viewing (Bradley et al., 1990; Vrana, Spence, & Lang, 1988). It has been suggested that, during US anticipation in conditioning tasks, general arousal contributes to startle magnitude (Dichter, Tomarken, & Baucom, 2002; Sabatinelli, Bradley, & Lang, 2001). This notion is supported by the finding that startle potentiation also occurs in

anticipation of pleasant US in appetitive conditioning (Bradley, Zlatar, & Lang, 2018; Mallan & Lipp, 2007; Mallan, Lipp, & Libera, 2008), whereas the startle reflex should be inhibited if it was valence-modulated (as it is during positive affective picture viewing Bradley et al., 1990; Hamm et al., 1993). Hence, both arousal during US anticipation as well as an acquired negative valence of the CS+ may contribute to the stimulus discrimination captured by the startle reflex.

Similarly, SCR may not only reflect cognitive threat appraisal as they have also been associated with reflexive fear circuitry involving the amygdala: Mangina and Beuzeron-Mangina (1996) demonstrated that amygdala stimulation elicits SCR in humans and Wood, Ver Hoef, and Knight (2014) found that both amygdala activity and SCR were similarly enhanced in response to noises presented during emotionally arousing pictures. Furthermore, both startle responses and onset SCR decline robustly during fear acquisition, which is often also reported for amygdala activity (Büchel, Morris, Dolan, & Friston, 1998; Leuchs et al., 2016; Lindner et al., 2015).

Therefore, both reflexive, subcortical processes and arousal during conscious threat appraisal may contribute to different physiological responses such as the startle reflex and SCR, but to a different extent. This may also depend on the chosen window of interest for the conditioned response: in particular SCR at stimulus onset may be more reflexive and amygdala-dependent, whereas SCR emerging at a later time point during stimulus presentations may reflect more anticipatory arousal (Cheng, Richards, & Helmstetter, 2007; Ohman, 1971, 1974).

4.2 Methodological considerations and limitations

Physiological outcome measures can have different advantages in different experimental settings (for an overview see Lonsdorf et al., 2017). In general, triggered responses like the startle reflex, the auditory pupil reflex or the pupillary light reflex have the advantage that they can be probed at any time throughout fear learning, also during neutral inter-trial intervals. Otherwise, continuous measures such as skin conductance or slow pupil dilations potentially offer information on tonic changes in arousal.

Pupillometry specifically may be advantageous in the MR environment and for longer tasks, since habituation, a comparatively low room temperature or constant background noise do not seem to affect phasic pupil dilations. However, pupil measurements are strongly influenced by light conditions and stimulus and background luminance have to be controlled. Moreover, data quality quickly deteriorates with strong gaze shifts that can result from visual exploration, which is why cues need to be presented focally. Such adaptations of the stimulus material can result in less naturalistic experimental conditions.

In our study, the fear acquisition phase contained a relatively high amount of stimulation (two different US as well as startle probes), which is likely associated with overall higher physiological activity. We have assured a stimulation-free period of (minimally) five seconds preceding each conditioned stimulus. However, slowly

recovering measures like skin conductance may still show cumulative effects from preceding stimulations. In similar vein, onset SCR cannot be disentangled from SCR in US-anticipation, given the relatively short stimulus durations in our study. Pupil responses at stimulus onset and offset were furthermore dominated by pupillary light reflexes, which future studies could prevent by using stimuli which are isoluminant to the background during inter-trial intervals (see for example Schneider et al., 2018). Furthermore, startle probes should have been adjusted to occur at the exact same time during CS+ and CS- stimulus presentations.

When combining different outcome measures, their interactions have to be considered. In particular startle probes may have influenced the fear learning process itself (Sjouwerman et al., 2016) and can be perceived as highly aversive: in our study, their aversiveness was comparable to the US, possibly due to their loudness. Despite balancing startle probes throughout task phases, they have caused systematic differences between conditions by eliciting fear-potentiated auditory pupil responses. Similarly, startle probes may have influenced SCR in a stimulus-specific manner. Nonetheless, we could with high certainty exclude the effects of the startle probes on onset SCR since an additional analysis, for which we defined a response window that lasted only until 0.6 s after startle probes, yielded highly similar results.

Longer stimulus presentations and applications of startle probes late throughout trials might reduce the interference of different processes (i.e. by temporally separating onset responses from anticipatory responses). However, startle responses still share the same window of interest with anticipatory pupil responses, as both are maximal in direct US anticipation (Grillon et al., 1993; Reinhard & Lachnit, 2002). Furthermore, longer trials would prolong the duration of the task, presumably leading to increased habituation in late task phases. Reducing the number of stimulus presentations could provide a solution to keep the original task duration, however, both options involve a loss of information. As different physiological readouts respond and recover on different temporal scales (in the range of milliseconds for the startle reflex and several seconds for SCR; see Lonsdorf et al., 2017 for an overview), simultaneous measurements demand a compromise for stimulus and inter-trial interval durations, the number of stimulus presentations and probe applications.

4.3 Conclusion

We simultaneously assessed skin conductance, startle EMG and pupil diameter during a two-day fear acquisition, extinction and recall paradigm. This was performed to provide a comparison of the different measures' properties and temporal dynamics throughout fear learning, and to relate pupillometric measures to other frequently used readouts of the conditioned response. Pupillometric measures, startle EMG and skin conductance each discriminated between CS+ and CS-, but correlated only weakly among each other. We found largest differences in response patterns during fear acquisition. Here, contrarily to pupillary measures, SCR and startle responses were strongly affected by habituation.

Our data suggest that pupil dilations roughly mirrored US expectancy ratings and increased in response to CS+ throughout fear acquisition. We found a fear-potentiation of the auditory pupil reflex to startle probes, which may constitute an interesting new readout for conditioned responding. We furthermore replicated findings of the negative association of pupil baseline diameter with superimposed phasic pupil dilations, emphasizing the importance of reporting – or accounting for – temporal dynamics in tonic pupil diameter. Physiological habituation was thereby not expressed in pupil responses, but rather in baseline pupil diameter.

Our data suggest that, despite showing some overlap, SCR, startle responses and pupillometric measures diverge as measures of the conditioned response. Simultaneous assessments of multiple physiological outcome measures of the conditioned response can hence offer complementary information about different cognitive-affective processes throughout fear learning.

5. References

Bacigalupo, F., & Luck, S. J. (2017). Event-related potential components as measures of aversive conditioning in humans. *Psychophysiology, 55*(4). https://doi.org/10.1111/psyp.13015

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology (2nd ed.)* (pp. 142-162). New York, NY, US: Cambridge University Press.

Bitsios, P., Philpott, A., Langley, R. W., Bradshaw, C. M., & Szabadi, E. (1999). Comparison of the effects of diazepam on the fear-potentiated startle reflex and the fear-inhibited light reflex in man. *J Psychopharmacol, 13*(3), 226-234. http://doi.org/10.1177/026988119901300303

Bitsios, P., Szabadi, E., & Bradshaw, C. M. (1996). The inhibition of the pupillary light reflex by the threat of an electric shock: a potential laboratory model of human anxiety. *J Psychopharmacol, 10*(4), 279-287. https://doi.org/10.1177/026988119601000404

Bitsios, P., Szabadi, E., & Bradshaw, C. M. (2004). The fear-inhibited light reflex: importance of the anticipation of an aversive event. *Int J Psychophysiol, 52*(1), 87-95. https://doi.org/10.1016/j.ijpsycho.2003.12.006

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology, 42*(1), 1-15. https://doi.org/10.1111/j.1469-8986.2005.00271.x

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1990). Startle reflex modification: emotion or attention? *Psychophysiology, 27*(5), 513-522. http://doi.org/10.1111/j.1469-8986.1990.tb01966.x

Bradley, M. M., Lang, P. J., & Cuthbert, B. N. (1993). Emotion, novelty, and the startle reflex: Habituation in humans. *Behavioral Neuroscience, 107*(6), 970-980. https://doi.org/10.1037/0735-7044.107.6.970

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607. http://doi.org/10.1111/j.1469-8986.2008.00654.x

Bradley, M. M., Zlatar, Z. Z., & Lang, P. J. (2018). Startle reflex modulation during threat of shock and "threat" of reward. *Psychophysiology, 55*(2). https://doi.org/10.1111/psyp.12989

Büchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain Systems Mediating Aversive Conditioning: an Event-Related fMRI Study. *Neuron, 20*(5), 947-957. http://doi.org/10.1016/S0896-6273(00)80476-6

Cheng, D. T., Richards, J., & Helmstetter, F. J. (2007). Activity in the human amygdala corresponds to early, rather than late period autonomic responses to a signal for shock. *Learn Mem, 14*(7), 485-490. https://doi.org/10.1101/lm.632007

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear Conditioning in Humans: The Influence of Awareness and Autonomic Arousal on Functional Neuroanatomy. *Neuron, 33*(4), 653-663. https://doi.org/10.1016/S0896-6273(02)00588-3

De Voogd, L. D., Fernández, G., & Hermans, E. J. (2016). Awake reactivation of emotional memory traces through hippocampal–neocortical interactions. *Neuroimage, 134*(1), 563-572. http://doi.org/10.1016/j.neuroimage.2016.04.026

Dichter, G. S., Tomarken, A. J., & Baucom, B. R. (2002). Startle modulation before, during and after exposure to emotional stimuli. *Int J Psychophysiol, 43*(2), 191-196. http://doi.org/10.1016/S0167-8760(01)00170-2

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Avila-Parcet, A., & Radua, J. (2015). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry, 21*(4). https://doi.org/10.1038/mp.2015.88

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci, 10*(2), 252-269. https://doi.org/10.3758/CABN.10.2.252

Grillon, C., & Ameli, R. (1998). Effects of threat and safety signals on startle during anticipation of aversive shocks, sounds, and airblasts. *Journal of Psychophysiology, 12*(4), 329-337.

Grillon, C., Ameli, R., Merikangas, K., Woods, S. W., & Davis, M. (1993). Measuring the time course of anticipatory anxiety using the fear-potentiated startle reflex. *Psychophysiology, 30*(4), 340-346. http://doi.org/10.1111/j.1469-8986.1993.tb02055.x

Grillon, C., & Baas, J. (2003). A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Clin Neurophysiol, 114*(9), 1557-1579. http://doi.org/10.1016/S1388-2457(03)00202-5

Hamm, A. O., Greenwald, M. K., Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1991). The fear potentiated startle effect. Blink reflex modulation as a result of classical aversive conditioning. *Integr Physiol Behav Sci, 26*(2), 119-126. http://doi.org/10.1007/BF02691034

Hamm, A. O., Greenwald, M. K., Bradley, M. M., & Lang, P. J. (1993). Emotional learning, hedonic change, and the startle probe. *J Abnorm Psychol, 102*(3), 453-465. http://doi.org/10.1037/0021-843X.102.3.453

Hamm, A. O., & Stark, R. (1993). Sensitization and aversive conditioning: effects on the startle reflex and electrodermal responding. *Integr Physiol Behav Sci, 28*(2), 171-176. https://doi.org/10.1007/BF02691223

Hamm, A. O., & Vaitl, D. (1996). Affective learning: awareness and aversion. *Psychophysiology, 33*(6), 698-710. http://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. http://doi.org/10.1016/j.ijpsycho.2005.01.006

Henderson, R. R., Bradley, M. M., & Lang, P. J. (2014). Modulation of the initial light reflex during affective picture viewing. *Psychophysiology, 51*(9), 815-818. http://doi.org/10.1111/psyp.12236

Hourdaki, E., Giakoumaki, S. G., Grinakis, V., Theou, K., Karataraki, M., & Bitsios, P. (2005). Parametric exploration of the fear-inhibited light reflex. *Psychophysiology, 42*(4), 447-455. https://doi.org/10.1111/j.1469-8986.2005.00301.x

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *J Cogn Neurosci, 23*(7), 1587-1596. https://doi.org/10.1162/jocn.2010.21548

Jovanovic, T., Keyes, M., Fiallos, A., Myers, K. M., Davis, M., & Duncan, E. J. (2005). Fear potentiation and fear inhibition in a human fear-potentiated startle paradigm. *Biol Psychiatry, 57*(12), 1559-1564. https://doi.org/10.1016/j.biopsych.2005.02.025

Jovanovic, T., Norrholm, S. D., Keyes, M., Fiallos, A., Jovanovic, S., Myers, K. M., . . . Duncan, E. J. (2006). Contingency awareness and fear inhibition in a human fear-potentiated startle paradigm. *Behav Neurosci, 120*(5), 995-1004. https://doi.org/10.1037/0735-7044.120.5.995

Kessler, R. C., Calabrese, J. R., Farley, P. A., Gruber, M. J., Jewell, M. A., Katon, W., . . . Wittchen, H. U. (2013). Composite International Diagnostic Interview screening scales for DSM-IV anxiety and mood disorders. *Psychol Med, 43*(8), 1625-1637. https://doi.org/10.1017/S0033291712002334

Koenig, S., Uengoer, M., & Lachnit, H. (2017a). Attentional Bias for Uncertain Cues of Shock in Human Fear Conditioning: Evidence for Attentional Learning Theory. *Front Hum Neurosci, 11*, 266. https://doi.org/10.3389/fnhum.2017.00266

Koenig, S., Uengoer, M., & Lachnit, H. (2017b). Pupil dilation indicates the coding of past prediction errors: Evidence for attentional learning theory. *Psychophysiology, 55*(4). https://doi.org/10.1111/psyp.13020

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychol Rev, 97*(3), 377-395. http://doi.org/10.1037/0033-295X.97.3.377

Lavin, C., San Martin, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Front Behav Neurosci, 7*, 218. https://doi.org/10.3389/fnbeh.2013.00218

Leuchs, L., Schneider, M., Czisch, M., & Spoormaker, V. I. (2016). Neural correlates of pupil dilation during human fear learning. *Neuroimage, 147*, 186-197. https://doi.org/10.1016/j.neuroimage.2016.11.072

Lindner, K., Neubert, J., Pfannmoller, J., Lotze, M., Hamm, A. O., & Wendt, J. (2015). Fear-potentiated startle processing in humans: Parallel fMRI and orbicularis EMG assessment during cue conditioning

and extinction. *Int J Psychophysiol, 98*(3 Pt 2), 535-545. https://doi.org/10.1016/j.ijpsycho.2015.02.025

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cogn Emot, 19*(2), 269-282. https://doi.org/10.1080/02699930441000319

Lipp, O. V., Sheridan, J., & Siddle, D. A. (1994). Human blink startle during aversive and nonaversive Pavlovian conditioning. *J Exp Psychol Anim Behav Process, 20*(4), 380-389. https://doi.org/10.1037/0097-7403.20.4.380

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., . . . Merz, C. J. (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

Lowenstein, O., Feinberg, R., & Loewenfeld, I. E. (1963). Pupillary Movements during Acute and Chronic Fatigue - a New Test for the Objective Evaluation of Tiredness. *Investigative Ophthalmology, 2*(2), 138-157.

Mallan, K. M., & Lipp, O. V. (2007). Does emotion modulate the blink reflex in human conditioning? Startle potentiation during pleasant and unpleasant cues in the picture-picture paradigm. *Psychophysiology, 44*(5), 737-748. https://doi.org/10.1111/j.1469-8986.2007.00541.x

Mallan, K. M., Lipp, O. V., & Libera, M. (2008). Affect, attention, or anticipatory arousal? Human blink startle modulation in forward and backward affective conditioning. *Int J Psychophysiol, 69*(1), 9-17. https://doi.org/10.1016/j.ijpsycho.2008.02.005

Mangina, C. A., & Beuzeron-Mangina, J. H. (1996). Direct electrical stimulation of specific human brain structures and bilateral electrodermal activity. *Int J Psychophysiol, 22*(1-2), 1-8. https://doi.org/10.1016/0167-8760(96)00022-0

Marois, A., Labonte, K., Parent, M., & Vachon, F. (2017). Eyes have ears: Indexing the orienting response to sound using pupillometry. *Int J Psychophysiol, 123*. https://doi.org/10.1016/j.ijpsycho.2017.09.016

Murphy, P. R., van Moort, M. L., & Nieuwenhuis, S. (2016). The Pupillary Orienting Response Predicts Adaptive Behavioral Adjustment after Errors. *PLoS One, 11*(3), e0151763. https://doi.org/10.1371/journal.pone.0151763

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology, 48*(2), 162-175. https://doi.org/10.1111/j.1469-8986.2010.01057.x

Ohman, A. (1971). Differentiation of conditioned and orienting response components in electrodermal conditioning. *Psychophysiology, 8*(1), 7-22.

Ohman, A. (1974). Orienting reactions, expectancy learning, and conditioned responses in electrodermal conditioning with different interstimulus intervals. *Biol Psychol, 1*(3), 189-200. https://doi.org/10.1016/0301-0511(74)90011-8

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *Int. J. Hum.-Comput. Stud., 59*(1-2), 185-198. https://doi.org/10.1016/s1071-5819(03)00017-x

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

Preuschoff, K., Hart, B. M., & Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Front Neurosci, 5*, 115. https://doi.org/10.3389/fnins.2011.00115

Purkis, H. M., & Lipp, O. V. (2001). Does Affective Learning Exist in the Absence of Contingency Awareness? *Learning and Motivation, 32*(1), 84-99. https://doi.org/10.1006/lmot.2000.1066

Reinhard, G., & Lachnit, H. (2002). Differential conditioning of anticipatory pupillary dilation responses in humans. *Biol Psychol, 60*(1), 51-68. https://doi.org/10.1016/S0301-0511(02)00011-X

Reinhard, G., Lachnit, H., & Konig, S. (2006). Tracking stimulus processing in Pavlovian pupillary conditioning. *Psychophysiology, 43*(1), 73-83. https://doi.org/10.1111/j.1469-8986.2006.00374.x

Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2001). Affective startle modulation in anticipation and perception. *Psychophysiology, 38*(4), 719-722. https://doi.org/10.1111/1469-8986.3840719

Schneider, M., Leuchs, L., Czisch, M., Samann, P. G., & Spoormaker, V. I. (2018). Disentangling reward anticipation with simultaneous pupillometry / fMRI. *Neuroimage, 178*. https://doi.org/10.1016/j.neuroimage.2018.04.078

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

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdiscip Rev Cogn Sci, 5*(6), 679-692. https://doi.org/10.1002/wcs.1323

Sjouwerman, R., Niehaus, J., Kuhn, M., & Lonsdorf, T. B. (2016). Don't startle me-Interference of startle probe presentations and intermittent ratings with fear acquisition. *Psychophysiology, 53*(12), 1889-1899. https://doi.org/10.1111/psyp.12761

Snowden, R. J., O'Farrell, K. R., Burley, D., Erichsen, J. T., Newton, N. V., & Gray, N. S. (2016). The pupil's response to affective pictures: Role of image duration, habituation, and viewing mode. *Psychophysiology, 53*(8), 1217-1223. https://doi.org/10.1111/psyp.12668

Society for Psychophysiological Research. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Tabbert, K., Merz, C. J., Klucken, T., Schweckendiek, J., Vaitl, D., Wolf, O. T., & Stark, R. (2010). Cortisol enhances neural differentiation during fear acquisition and extinction in contingency aware young women. *Neurobiol Learn Mem, 94*(3), 392-401. https://doi.org/10.1016/j.nlm.2010.08.006

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

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon Bull Rev*. https://doi.org/10.3758/s13423-018-1432-y

Visser, R. M., Kunze, A. E., Westhoff, B., Scholte, H. S., & Kindt, M. (2015). Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. *Psychoneuroendocrinology, 55*, 8-20. https://doi.org/10.1016/j.psyneuen.2015.01.021

Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nat Neurosci, 16*(4), 388-390. https://doi.org/10.1038/nn.3345

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: a new measure of emotion? *J Abnorm Psychol, 97*(4), 487-491. https://doi.org/10.1037/0021-843X.97.4.487

Widmann, A., Schroger, E., & Wetzel, N. (2018). Emotion lies in the eye of the listener: Emotional arousal to novel sounds is reflected in the sympathetic contribution to the pupil dilation response and the P3. *Biol Psychol, 133*, 10-17. https://doi.org/10.1016/j.biopsycho.2018.01.010

Wilhelm, H., Ludtke, H., & Wilhelm, B. (1998). Pupillographic sleepiness testing in hypersomniacs and normals. *Graefes Archive for Clinical and Experimental Ophthalmology, 236*(10), 725-729. https://doi.org/10.1007/s004170050149

Wood, K. H., Ver Hoef, L. W., & Knight, D. C. (2014). The amygdala mediates the emotional modulation of threat-elicited skin conductance response. *Emotion, 14*(4), 693-700. https://doi.org/10.1037/a0036636

6. Author Notes

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med Programme (grant # 01ZX1314J). The authors would like to thank Tanja Jovanovic and Seth Norrholm for their technical support in the laboratory setup. Further thank goes to the whole “Biological classification of mental disorders” (BeCOME) study team for their support in participant recruitment, as well as to Dorothee Pöhlchen Miriam Kraft and Florian Binder for their support in data acquisition and proofreading.

Name and address for reprints

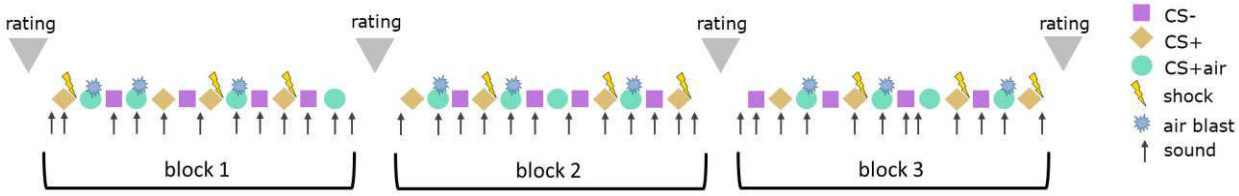
Dr. Victor Spoormaker, Max Planck Institute of Psychiatry, Department of Translational Research in Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany. Telephone: +49 (0) 89-30622-392 Email: spoormaker@psych.mpg.de.

Supplementary Material

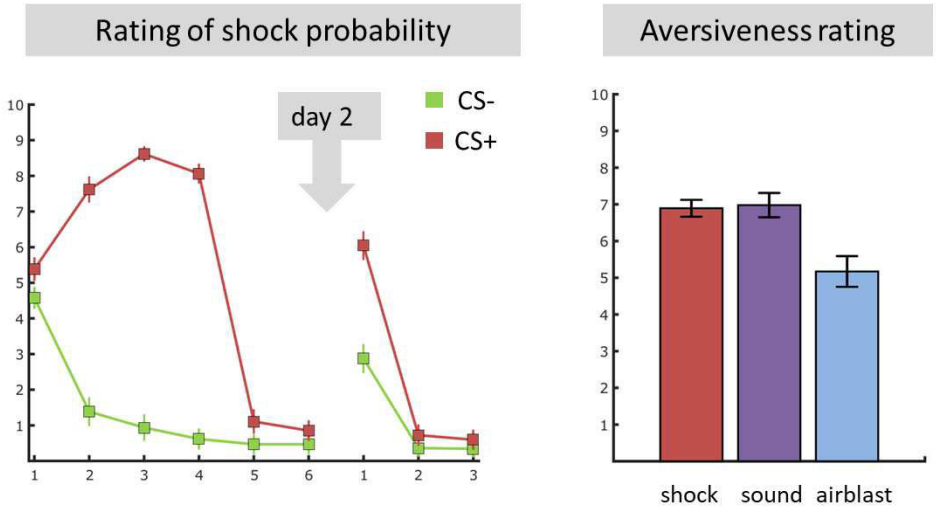
Supplementary Information: light conditions and stimulus color-codes

The approximate luminance in the laboratory was 300 Lux when facing the screen from the headrest (measured with a Luxmeter Testoterm 0500, Lenzkirch, Germany). The stimulus color-codes were [191, 119, 221], [120, 221, 189] and [220, 190, 120], each referring to a brightness of 160. Stimuli filled approximately 8.5 cm² on the screen (screen size: 30 × 37.5 cm) and the fixation cross had a line length of 1.5 cm with a thickness of 0.3 cm on the screen. Stimuli and fixation cross were presented centrally on a black background, resulting in an average screen luminance of 13.8 and 0.25, respectively.

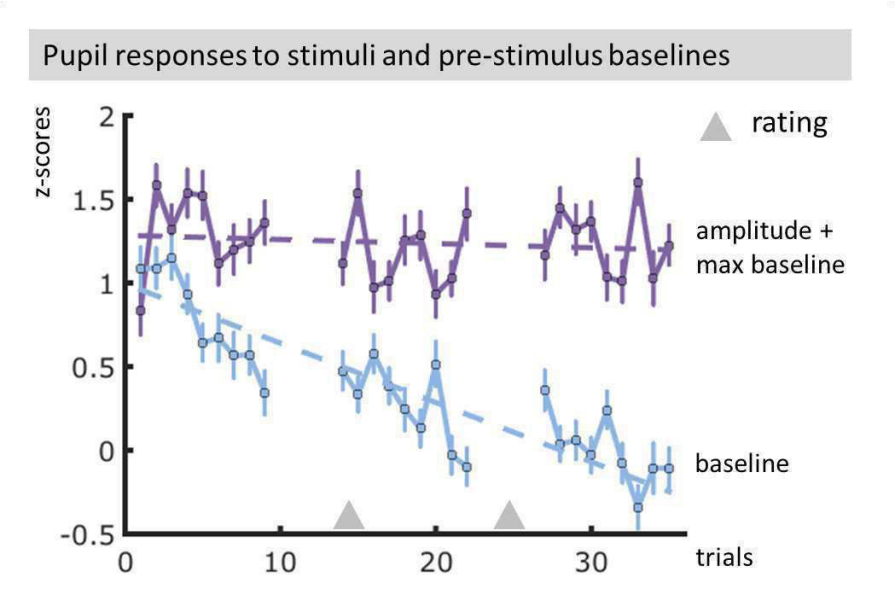
1. Supplementary Figures



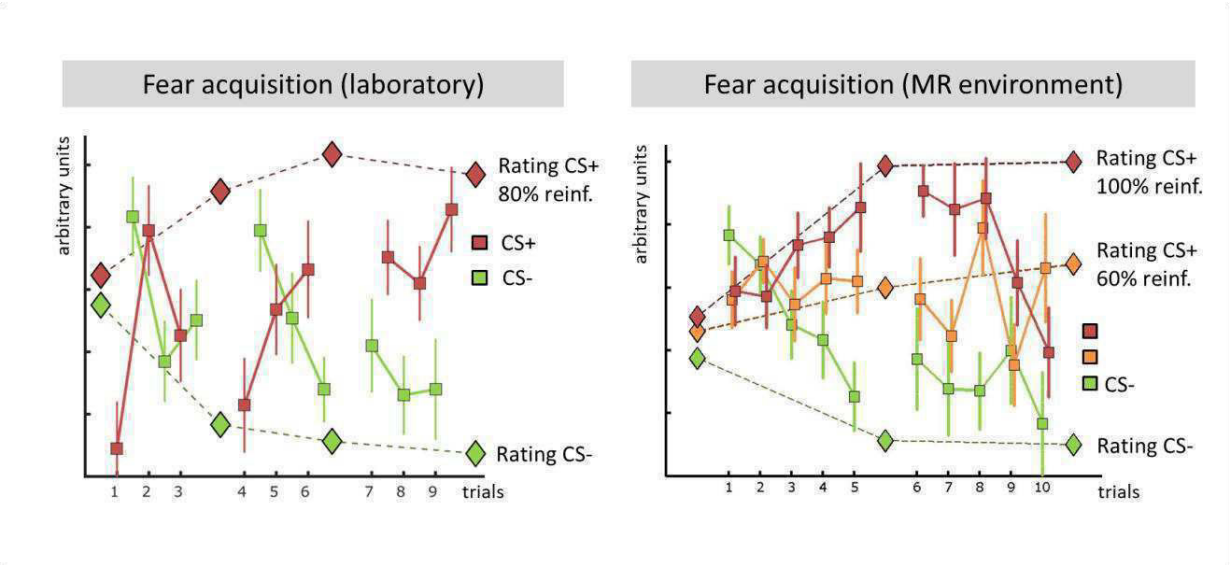
Supplementary Figure S1. Stimulus order during fear acquisition. The three conditioned stimuli (one CS- and two CS+) were presented in a pseudo-randomized order with intermittent ratings of US-probability. During fear acquisition, 75% of CS+ trials were coupled with their respective US. Startle probes occurred equally often during all stimulus types (in 75% of trials) and during inter-trial intervals. The amount of stimulus presentations, startle probes and US applications were balanced between the ratings of US probability.



Supplementary Figure S2. Left: subjective ratings of shock probability during fear acquisition and recall on the first and second test day, respectively. Right: subjective ratings of stimulus aversiveness. Vertical bars: SEM.



Supplementary Figure S3. Relationship of pre-stimulus baseline pupil diameter and succeeding slow pupil dilations during fear acquisition (for all stimulus presentations of CS-, CS+ and CS+air, including startle probes). For illustration purposes, the maximal baseline value was added to all response amplitudes (both reflect z-scores across trials from both test days which are centered around zero). Dashed lines: linear regression; vertical lines: SEM.



Supplementary Figure S4. Ratings of US probability in relation to pupil responses. Left: sample of the current study, $N = 47$; Right: sample of $N = 28$ subjects from Leuchs et al. (2016), pupil responses assessed during functional magnetic resonsnce imaging (fMRI). Vertical lines: SEM; Diamonds: ratings in arbitrary units.

2. Supplementary Tables

Supplementary Table S1. Stimulus effect and temporal analysis for all outcome measures: results of paired t-tests (CS+air > CS-), estimates of the stimulus effect (Cohen's *d*) and results of rmANOVAs for ratings of airblast probability and block-wise averages of physiological responses; rmANOVAs contained the factors stimulus (CS- & CS+air) and time (three values for fear acquisition, two values for extinction and two values for recall). Note that only the other CS+ (reinforced by shocks) underwent extinction.

	t-test CS+ > CS-	Cohen's d	Stimulus	Time	Stimulus × Time
Ratings of airblast probability					
Fear acquisition	t (45) = 19.06, p < .001	2.78	F (1,46) = 312.83, p < . 001	F (3,138) = 2.12, p < .05	F(3,138) = 112.41, p < . 001
Extinction	t (45) = 3.91, p < .001	.57	F (1,46) = 15.45, p < . 001	F (3,138) = 0.00, p = .95	F(3,138) = 0.18, p = .67
recall	t (45) = 5.72, p = .29	.83	F (1,46) = 32.71, p < . 001	F (3,138) = 134.01, p < .001	F(3,138) = 25.81, p < . 001
SCR					
Fear acquisition	t (45) = 0.59, p = .29	.07	F (1,42) = 0.17, p = .68	F (2,84) = 12.90, p < .001	F (2,84) = 0.11, p = .90
recall	t (45) = 1.02, p = .31	.18	F (1,46) = 0.05, p = .82	F (1,46) = 22.08, p < .001	F (1,46) = 3.76, p = .06
Startle response					
Fear acquisition	t (46) = 4.63, p < .001	.67	F (1,46) = 21.39, p < .001	F (2,92) = 67.37, p < .001	F (2,92) = 0.73, p = .48
recall	t (46) = 4.90, p < .001	.71	F (1,46) = 23.97, p < .001	F (1,46) = 131.09, p < .001	F (1,46) = 0.74, p = .39
Slow pupil dilations					
Fear acquisition	t (46) = 0.14, p = .89	.02	F (1,42) = 0.27, p = .61	F (2,84) = 2.60, p = .08	F (2,84) = 3.67, p < .05
recall	t (46) = 2.79, p < .005	.41	F (1,46) = 7.57, p < .01	F (1,46) = 13.78, p < .005	F (1,46) = 0.28, p = .60
Auditory pupil reflex to the startle probe					
Fear acquisition	t (46) = 2.22, p < .05	.32	F (1,42) = 5.37, p < .05	F (2,84) = 1.90, p = .16	F (2,84) = 0.95, p = .39
recall	t (46) = 0.38, p = .35	.06	F (1,46) = 0.10, p = .75	F (1,46) = 9.36, p < .005	F (1,46) = 0.01, p = .91

Supplementary Table S2: Comparisons of physiological responses to the different CS+ (one followed by shocks, one followed by airblasts). Results of paired t-tests for fear acquisition and recall. Note that only the CS+ which was reinforced with shocks underwent extinction.

	t-test CS+ versus CS+air
SCR: larger responses to CS+ followed by shocks	
Fear acquisition	t (46) = 2.44, p < .05
Recall	t (46) = 1.98, p = .05
Startle: larger responses to CS+ followed by airblasts	
Fear acquisition	t (46) = 7.4, p < .001
Recall	t (46) = 4.85, p < .001
Slow pupil dilations: no difference	
Fear acquisition	t (46) = 0.92, p = .36
Recall	t (46) = 0.12, p = .91
Auditory pupil reflex to the startle probe: no difference	
Fear acquisition	t (46) = 0.20, p = .84
Recall	t (46) = 0.66, p = .51

Supplementary Table S3. Main effects of measure, stimulus and time from two rmANOVAs including SCR, startle responses and one pupillary measure, respectively. Factors include measure (SCR, startle responses and either slow pupil dilations or auditory pupil reflexes to the startle probe), stimulus (block-wise averages of responses to CS- and CS+) and time (three values for fear acquisition, two values for extinction and two values for recall).

	Measure	**Stimulus**	**Time**
Slow pupil dilations			
Fear acquisition	F (2,80) = 5.89, p < .005	F (1,40) = 17.82, p < .001	F (2,80) = 43.49, p < . 001
Extinction	F (2,84) = 13.54, p < .001	F (1,42) = 0.65, p = .42	F (1,42) = 5.72, p < .05
recall	F (2,82) = 3.154, p < .05	F (1,41) = 10.84, p < .005	F (1,41) = 70.03, p < .001
Auditory pupil reflex to the startle probe			
Fear acquisition	F (2,76) = 3.09, p < .05	F (1,38) = 26.17, p < .001	F (2,76) = 33.79, p < . 001
Extinction	F (2,80) = 27.37, p < .001	F (1,40) = 1.43, p = .24	F (1,40) = 18.42, p < .001
recall	F (2,78) = .80, p = .45	F (1,39) = 3.92, p = .06	F (1,39) = 65.00, p < .001

Supplementary Table S4. Interactions of measure, stimulus and time from two rmANOVAs including SCR, startle responses and one pupillary measure, respectively. Factors include measure (SCR, startle responses and either slow pupil dilations or auditory pupil reflexes to the startle probe), stimulus (block-wise averages of responses to CS- and CS+) and time (three values for fear acquisition, two values for extinction and two values for recall).

	Measure × Stimulus	**Measure × Time**	**Stimulus × Time**	**Measure × Stimulus × Time**
Slow pupil dilations				
Fear acquisition	$F(2,80) = 8.32$, $p < .001$	$F(4,160) = 8.32$, $p < .001$	$F(2,80) = 1.82$, $p = .17$	$F(4,160) = 3.83$, $p < .01$
Extinction	$F(2,84) = 0.25$, $p = .78$	$F(2,84) = 0.94$, $p = .40$	$F(1,42) = 0.26$, $p = .61$	$F(2,84) = 2.58$ $p = .08$
recall	$F(2,82) = 2.23$, $p = .12$	$F(2,82) = 2.38$, $p = .10$	$F(1,41) = 2.03$, $p = .16$	$F(2,84) = 0.44$, $p = .65$
Auditory pupil reflex to the startle probe				
Fear acquisition	$F(2,76) = 4.56$, $p < .05$	$F(4,152) = 13.36$, $p < .001$	$F(2,76) = 0.44$, $p = .65$	$F(4,152) = 4.29$, $p < .005$
Extinction	$F(2,80) = 0.02$, $p = .99$	$F(2,80) = 0.70$, $p = .50$	$F(1,40) = 0.30$, $p = .59$	$F(2,80) = 1.12$ $p = .308$
recall	$F(2,78) = 4.24$, $p < .05$	$F(2,78) = 6.50$, $p < .005$	$F(1,39) = 0.91$, $p = .34$	$F(2,78) = 1.17$, $p = .32$

3. Neural correlates of pupil dilation during human fear learning

3.1 Summary

In the second study, pupillometry was simultaneously acquired with fMRI measurements during a fear acquisition and extinction task. Pupil dilations were larger in response to CS+ than CS-, with strongest stimulus discrimination shortly before US onset. Pupil dilations furthermore discriminated between two CS+ with different reinforcement rates: responses to a fully reinforced CS+ initially increased more strongly than responses to an intermittently reinforced CS+ (60% reinforcement rate), but declined later throughout fear acquisition (see Figure 2 A). Responses to the intermittently reinforced CS+ remained elevated during late fear learning. Based on these observations, we propose that both threat appraisal and uncertainty about reinforcement contingencies contributed to the magnitude of pupil responses during fear learning. The discrimination between CS+ and CS- was detectable until the end of extinction, which suggests that pupillometry is a highly sensitive readout of conditioned fear that does not seem to be affected by physiological habituation. During fear acquisition, we found the expected brain regions of the 'fear' or salience network (as reported by Fullana et al., 2015) to be associated with the CS+ > CS- contrast, while the reverse contrast yielded activity in the vmPFC, among others. We correlated the magnitude of slow pupil dilations in response to CS+ and CS- to concurrent brain activation (approximated by the blood-oxygen level dependent [BOLD] signal measured with fMRI). The magnitude of pupil dilations was positively correlated with clusters largely overlapping with the salience network, including dACC, bilateral insula and thalamus. Pupil response magnitude was negatively correlated with activity in vmPFC. When controlling these correlations for the CS+ > CS- stimulus contrast (i.e. searching for activity associated with pupil dilations within instead of across stimulus types), we still found a significant association of dACC activity with pupil response magnitude at the trial-level. These results show that slow pupil dilations during fear learning reflect activity in brain regions commonly involved in threat (salience network) and safety signaling (vmPFC). Our data therefore suggests a stimulus-unspecific association of phasic pupil responses with dACC activity, specifically.

3.2 Declaration of author contributions

The study was conceived and designed by Victor Spoormaker (VS) and Laura Leuchs (LL). LL programmed the experiment. VS, LL and Max Schneider (MS) implemented the experiment in the MR environment with help from Brice Fernandez (BF). LL recruited the participants with help from Ines Eidner (IE). LL and IE carried out the experiments with help from MS and VS. VS, LL, MS and Michael Czisch (MC) contributed to the interpretation of the results. LL performed the data analysis with help from VS and BF. LL took the lead in writing the manuscript with support from VS. All authors provided critical feedback to the manuscript.

NeuroImage 147 (2017) 186–197

Contents lists available at ScienceDirect

NeuroImage

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

Neural correlates of pupil dilation during human fear learning

Laura Leuchs, Max Schneider, Michael Czisch, Victor I. Spoormaker*

Max Planck Institute of Psychiatry, 80804 Munich, Germany

ARTICLE INFO

Keywords:
Pupillometry
Fear conditioning
Fear extinction
fMRI
Dorsal anterior cingulate cortex
Salience network

ABSTRACT

Background: Fear conditioning and extinction are prevailing experimental and etiological models for normal and pathological anxiety. Pupil dilations in response to conditioned stimuli are increasingly used as a robust psychophysiological readout of fear learning, but their neural correlates remain unknown. We aimed at identifying the neural correlates of pupil responses to threat and safety cues during a fear learning task.

Methods: Thirty-four healthy subjects underwent a fear conditioning and extinction paradigm with simultaneous functional magnetic resonance imaging (fMRI) and pupillometry. After a stringent preprocessing and artifact rejection procedure, trial-wise pupil responses to threat and safety cues were entered as parametric modulations to the fMRI general linear models.

Results: Trial-wise magnitude of pupil responses to both conditioned and safety stimuli correlated positively with activity in dorsal anterior cingulate cortex (dACC), thalamus, supramarginal gyrus and insula for the entire fear learning task, and with activity in the dACC during the fear conditioning phase in particular. Phasic pupil responses did not show habituation, but were negatively correlated with tonic baseline pupil diameter, which decreased during the task. Correcting phasic pupil responses for the tonic baseline pupil diameter revealed thalamic activity, which was also observed in an analysis employing a linear (declining) time modulation.

Conclusion: Pupil dilations during fear conditioning and extinction provide useful readouts to track fear learning on a trial-by-trial level, particularly with simultaneous fMRI. Whereas phasic pupil responses reflect activity in brain regions involved in fear learning and threat appraisal, most prominently in dACC, tonic changes in pupil diameter may reflect changes in general arousal.

1. Introduction

Fear conditioning and extinction paradigms can be used to assess individual differences in fear learning and provide useful etiological models for anxiety disorders. Previous studies have revealed different response patterns during fear conditioning and extinction in healthy controls and psychiatric patients suffering from post-traumatic stress disorder (PTSD; Blechert et al., 2007; Glover et al., 2011; Grillon et al., 2009; Pole, 2007), panic disorder (Grillon et al., 2008; Lissek et al., 2010), social anxiety disorder (Lissek et al., 2008b) and others. A meta-analysis by Lissek et al. (2005b) and a recent update by Duits et al. (2015) reported increased physiological responding to safety cues (CS-) during fear acquisition and increased responses to fear conditioned stimuli (CS+) during extinction in patients suffering from anxiety disorders and PTSD.

To objectively assess fear learning, autonomous arousal in response to CS- as compared to CS+ is measured. This can be done by contrasting skin conductance responses (SCR), which are mainly determined by changes in sympathetic activation (Dawson et al., 2007). Another commonly applied method is startle electromyography (EMG), during which the startle reflex to sudden loud noises is quantified (for a review of startle responsivity in clinical populations see Vaidyanathan et al., 2009). The startle reflex is modulated by emotional valence (Grillon and Baas, 2003) and yields stronger responses to noises applied during CS+ presentations as compared to CS- presentations (Lindner et al., 2015; Lipp et al., 1994; Van Well et al., 2012).

However, both SCR and startle EMG have some disadvantages. SCRs can occur spontaneously and show a considerable amount of inter- and intra-subject variability (Bach et al., 2009; Benedek and Kaernbach, 2010), leading to rather high levels of noise. Furthermore, SCR shows quick habituation, leading to a high amount of zero responses (e.g., defined as responses below 0.01 mS, Dawson et al., 2007) and generally to weak-to-moderate effect sizes for the discrimination of CS- and CS+ (Pineles et al., 2009). The fear potentiated startle has been shown to yield strong effects sizes (Lissek et al., 2008a), but this measure also shows strong habituation effects (Grillon and Baas, 2003) and is more difficult to apply during fMRI measure-

* Correspondence to: Max Planck Institute of Psychiatry, Department of Translational Research in Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany.
E-mail address: spoormaker@psych.mpg.de (V.I. Spoormaker).

http://dx.doi.org/10.1016/j.neuroimage.2016.11.072
Received 11 August 2016; Accepted 30 November 2016
Available online 01 December 2016
1053-8119/

ments due to the high acoustic background noise (but see Lindner et al., 2015 for an example of combined startle EMG and fMRI). Moreover, the startle sounds themselves can elicit SCRs and can be perceived to be aversive (Lissek et al., 2005a), potentially interfering with cognitive processes of interest.

Pupillometry offers a promising, complementary method for the quantification of the conditioned response. It is non-aversive, not interfering with other measurements and can be combined with functional magnetic resonance imaging (fMRI). The pupil diameter is determined by the sympathetically innervated dilator muscle and the parasympathetically innervated sphincter muscle (Beatty and Lucero-Wagoner, 2000). It is primarily influenced by optical reflexes for light and distance, however, tonic fluctuations and phasic dilations of the pupil associated with mental processes can be detected (Beatty and Lucero-Wagoner, 2000). While tonic changes in pupil diameter have mainly been related to general alertness (Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010) and wakefulness (Lowenstein et al., 1963; Wilhelm et al., 1998), phasic pupil dilations are sensitive to various cognitive and affective manipulations (for a review see Sirois and Brisson, 2014). The pupil dilates in response to aversive (Wiemer et al., 2014) and emotionally arousing stimuli (Bradley et al., 2008; Partala and Surakka, 2003) and can hence be used as a readout of fear-related processes.

Reinhard and Lachnit (2002) were the first to use pupil dilation during CS presentation as a physiological readout of human fear conditioning and reported stronger pupil dilations in response to CS + than to CS- (Reinhard and Lachnit, 2002; Reinhard et al., 2006). In the following years, pupil dilations have been used as readout of fear conditioning in several studies (De Voogd et al., 2016a; De Voogd et al., 2016b; Morriss et al., 2015; Visser et al., 2016; Visser et al., 2015; Visser et al., 2013). Furthermore, there is evidence for pupillary constriction during presentations of CS- (Pollak et al., 2010), suggesting that the pupil size may also reflect processes of fear inhibition.

The method for quantifying the pupillary response varies across studies: while a subtraction of the pre-stimulus baseline pupil diameter is common (Reinhard and Lachnit, 2002; Reinhard et al., 2006; Visser et al., 2016, 2015, 2013), the response has been defined as the mean pupil diameter spanning the whole CS presentation window (Morriss et al., 2015; Pollak et al., 2010) or as the peak change in pupil diameter occurring during CS presentation (Visser et al., 2016; Visser et al., 2013). Reinhard and Lachnit (2002) demonstrated that pupil dilations discriminate most strongly between CS- and CS+ in a time window immediately preceding the unconditioned stimulus (US). Visser et al. (2015) found that the baseline-to-peak difference during CS presentation yielded identical results as the change from baseline to the second preceding US onset, indicating that the peak pupil response occurs shortly before US onset.

Besides dilating to threat stimuli, the reflexive constriction of the pupil in response to light flashes is partly attenuated during the anticipation of aversive stimuli (Bitsios et al., 1996, 2004; Hourdaki et al., 2005). Comparable to the fear potentiated startle, this 'fear inhibited light reflex' to light flashes may be modulated by emotional valence (Bitsios et al., 2004). Reinhard et al. (2006) observed such an initial inhibition of the light reflex to CS+, albeit with smaller effect sizes than pupil dilation.

Pupillary responses to CS- and CS+ have been assessed with simultaneous fMRI, mainly as indicator of conditioned responding and fear recall (De Voogd et al., 2016a, 2016b; Morriss et al., 2015; Visser et al., 2016, 2015, 2013). To our knowledge, the neural correlates of pupil dilations during fear learning have not yet been reported. In this study, we aim to identify activity associated with pupil responses to CS- and CS+ during a fear learning task comprising fear conditioning and extinction. If the magnitude of trial-wise pupil dilations correlates with brain regions involved in fear expression or threat appraisal, this would provide support for the notion that pupil dilations are a meaningful trial-by-trial readout of the conditioned response.

In fMRI experiments, the comparison CS+ > CS- typically yields a characteristic pattern of activation, referred to as the fear network (Etkin and Wager, 2007; Fullana et al., 2015). This network comprises the dorsal anterior cingulate cortex (dACC), bilateral anterior insula, thalamus and parts of the striatum, among others; amygdala activation has been specifically associated with early phases of fear conditioning in some studies (for a meta-analysis and a review see Fullana et al., 2015; Sehlmeyer et al., 2009). The reverse contrast CS- > CS+ has been associated with activity in ventromedial prefrontal cortex (vmPFC; Milad and Quirk, 2012), lateral orbitofrontal cortex (OFC), hippocampus and posterior cingulate cortex (PCC; Fullana et al., 2015).

As there are no direct cortical inputs to the pupillary dilator or constrictor muscles (Beatty and Lucero-Wagoner, 2000), cognitive influences on pupil diameter must be conveyed in an indirect way. Activity of the locus coeruleus (LC) in the brainstem has been found to strongly correlate with pupil diameter in the monkey (Aston-Jones and Cohen, 2005; Joshi et al., 2016) and also in the human brain (Gilzenrat et al., 2010; Murphy et al., 2014; Sterpenich et al., 2006). As LC receives direct inputs from anterior cingulate cortex (ACC) and OFC in the monkey brain (Aston-Jones and Cohen, 2005), it may relay activity from these cortical regions into autonomous arousal and related pupil dilations. There is also a two-way excitatory connection between LC and amygdala (Samuels and Szabadi, 2008), which have been found to co-activate in fear-related processes (Liddell et al., 2005; Sears et al., 2013). Given these anatomical connections, we may expect ACC, orbitofrontal cortex and amygdala activity to be associated with the magnitude of pupil dilations during fear learning.

In addition to these anatomical considerations, we would expect pupil dilations in a fear learning task to be associated with regions of the fear network (especially dACC and amygdala). As the pupil is known to dilate more strongly in response to CS+ than to CS-, the neural correlates of pupil dilations should resemble the CS+ > CS- contrast. Reversely, smaller pupil responses may reflect fear inhibition and the activity pattern associated with the CS- > CS+ contrast (e.g., medial prefrontal cortex).

Previous work has shown that most robust differences in pupil responses to CS- and CS+ are found in temporal proximity to the US (Reinhard et al., 2006). We therefore focused on the change in pupil diameter from CS onset until shortly before US administration as a trial-wise readout of the conditioned response, but we also aimed to evaluate the initial light reflex at stimulus onset. Furthermore we explored the dynamics of tonic changes in pupil diameter during the fear learning task and investigated its relationship to phasic pupil responses. To further assess potential effects of uncertainty and expectancy on pupil dilations during fear learning, we used two CS+ with different reinforcement rates: higher reinforcement rates may provoke larger pupil dilations during fear conditioning due to higher threat appraisal, yet partial reinforcement may provoke larger pupil dilations due to uncertainty. Finally, we evaluated whether pupil size dynamics also reflect the so-called partial reinforcement extinction effect (PREE), which has been demonstrated for SCR (Grady et al., 2016). In this context, we would expect stronger pupillary – and behavioral responses – to a partially reinforced CS+ during extinction.

2. Material and methods

2.1. Participants

Thirty-four healthy subjects (mean [M] age=25.6, standard deviation [SD]=3.0, 18 male) participated in a fear learning task with simultaneous fMRI and pupillometry recordings. All subjects were right-handed, non-smokers and had normal or (contact lens) corrected vision. Prior to participation, subjects underwent an interview and a clinical MRI screening to exclude participants with present or past psychiatric or neurological disorders, or current use of psychotropic medication. The study protocol was in accordance with the Declaration

of Helsinki and was approved by a local ethics committee. After explanation of the study protocol, subjects gave written and informed consent and were reimbursed for participation.

2.2. Fear learning task and experimental procedure

The entire fear learning task comprised fear conditioning and immediate (unannounced) extinction. The three conditioned stimuli (one CS- and two CS+) consisted of three squares in different colors of equal size and brightness (colors were counter-balanced across subjects). All stimuli were presented using Presentation Software (Neurobehavioral Systems Inc., Berkeley, California, USA) and were displayed in the middle of a monitor, which was located at the end of the scanner bore. The full screen could be seen by the subjects through a first surface mirror, which was attached to the scanner head coil. Stimuli occupied 2.3°×2.3° of the visual angle and were presented on a black background.

The US consisted of a mild electric shock, which was administered to the back of the right wrist through gold electrodes (Digitimer Stimulator, Digitimer Ltd., Hertfordshire, UK). Shock strength was calibrated for each subject individually by starting at 2 mA stimulation and increasing the stimulus intensity in 1 mA steps until the shock was perceived as uncomfortable, but not yet painful (M=15.8 mA, SD=9.1 mA).

Stimuli were presented in a pseudo-randomized order for 4 s each, with no more than two subsequent trials of the same stimulus type. The two CS+ were associated with the US, one of them being reinforced in 60% (CS60+), and the other in 100% (CS100+) of the trials during the fear conditioning phase. In reinforced trials, shocks were administered 3.5 s after stimulus onset. The third stimulus was never followed by a US and hence served as CS-. Fear conditioning comprised 10 presentations of CS-, CS60+ and CS100+, respectively. For subsequent extinction, participants were randomized into two groups: 8 trials of either CS60+ (19 subjects) or CS100+ (15 subjects) were presented interspersed with 8 trials of CS-. Shocks were no longer applied during these trials, resulting in extinction of one of the previously conditioned stimuli (from this point on referred to as CSext). All trials were separated by inter-trial intervals of varying length (average duration of 12 ± 2 s), during which a white fixation cross was presented in the middle of the screen. The entire fear learning task (conditioning and extinction) comprised 46 trials and lasted 16 min and 24 s.

Subjective ratings of shock expectancy were assessed before and after the task, as well as at intermediate time points: halfway through the fear conditioning phase (after 5 trials of each stimulus, 15 trials in total), at the end of fear conditioning (after 10 trials of each stimulus, 30 trials in total) and halfway through the extinction phase (after 4 additional trials of CS- and CSext, 38 trials in total). Participants were asked to rate their shock expectancy within 10 s for each stimulus separately on a scale from 0% to 100% (in steps of 10%) by pressing a left and right button on a response device. Prior to the fear learning task, subjects were familiarized with the fMRI sequence and the stimuli in a brief habituation phase, with 6 non-reinforced stimulus presentations and one rating per stimulus to practice the procedure. After habituation, subjects were not explicitly informed about CS-US contingencies, but were told that the visual stimuli might be followed by electric shocks. They were asked to give an estimation of the shock probability for each stimulus separately, whenever the subjective rating appeared on the screen. Instructions were not repeated after the fear learning task had started and conditioning trials were immediately followed by extinction trials without interruption or any change of context.

2.3. Pupillometry

2.3.1. Data acquisition and preprocessing

The eye tracking system (Eyelink 1000 Plus, SR Research, Osgoode, ON, Canada) was located at the end of the scanner bore and tracked pupil diameter and gaze of the right eye via a mirror attached to the head coil (sampling rate: 250 Hz). A standard nine-point calibration procedure was conducted to locate the gaze position on the screen. Pupil data processing was performed in Matlab (version 2015a, MathWorks, Natick, USA). First, eye blinks were replaced by linear interpolation over the missing data points (Beatty and Lucero-Wagoner, 2000). Eye-blink related artifacts, consisting of pupil loss and re-detection, differed slightly across individuals. To remove artifacts while keeping data loss at a minimum, we semi-automatically adapted the interpolation window from shortly before (M=56 ms, range=20–60 ms, SD=8 ms) to shortly after eye blinks (M=128 ms, range=40–180 ms, SD=28 ms). Second, to control for variability across subjects, the entire pupil data per subject (recorded in arbitrary units) were z-transformed and all further analyses were carried out with the pupil diameter expressed in z-scores. Pupil data were then smoothed with a 200 ms sliding window and segmented per trial, spanning 0.5 s pre-stimulus (used as baseline) and the subsequent 4 s of stimulus presentation. Finally, invalid trials were identified by inspecting all segments for three exclusion criteria: first, trials containing more than 50% of interpolated data points were discarded (M=0.9% of trials per subject, SD=1.4%; see also Siegle et al., 2003; Visser et al., 2016, 2013). Second, we excluded trials with unnaturally sudden and large jumps in pupil diameter, which are typically caused by undetected blink artifacts or sudden changes in gaze direction. Because the pupillary light reflex to the stimulus caused a steep initial decrease in pupil diameter at each trial, we first split the trial segments into epochs: a 0.5 s epoch for the baseline and 4 epochs of each 1 s for the subsequent stimulus presentation. Then the first order derivatives were examined for every single epoch. If the derivative of an epoch differed by more than 3.3 standard deviations from the mean derivative of the same epoch of all other trials in this subject, the whole corresponding trial was discarded as an outlier (see Kafkas and Montaldi, 2011). In this way, only trials with exceptionally steep and sudden increases or decreases in pupil diameter (which could not be explained by light accommodation) were excluded (M=2.5% of trials per subject, SD=0.9%). Third, as strong shifts in gaze can impair pupil detection, we excluded trials in which subjects did not focus on the center of the screen for over 0.5 s during the trial. For this purpose, the average gaze shift during trials was determined across all subjects (1.81°×1.84° of visual angle), and a window of 3.3 standard deviations of the average gaze shift was defined around the center of the screen (corresponding to a visual angle of 6.0°×6.1°). Trials in which the subjects' gaze dwelled outside this window for longer than 0.5 s (i.e., trials containing artifacts affecting more than 0.5 s, comparable to procedures applied by Reinhard et al., 2006; Visser et al., 2013) were discarded from further analysis (M=8.9% trials per subject, SD=5.4%).

These three exclusion criteria partly classified the same segments as invalid, resulting in an average of 5 invalid trials out of 46 per subject (corresponding to 11% on average, SD=5.7%). There was no difference in the amount of missing trials across the three stimulus types ($F_{(2,50)}$=1.70, p=.194). Invalid trials were treated as missing data for further analysis. To assure that only attentive subjects with good data quality were included, subjects with over 20% of invalid trials were excluded from further analysis. This left 28 data sets for separate analyses of fear conditioning and extinction phases (partly different subjects) and a total of 26 data sets for analyses of the entire fear learning task.

As we anticipated that the discrimination in response to CS- and CS + would be maximal in close temporal proximity to US onset, we defined the last 0.5 s before shock administration as the time window of interest. The pupil response was calculated for each trial by subtracting the pre-stimulus baseline (average over the 0.5 s before stimulus onset) from the maximum pupil dilation between 3 and 3.5 s after stimulus onset (just before US application). The procedure of determining the pupil responses is also illustrated in Fig. 1A and is based on the method of calculating the peak pupil dilation as proposed

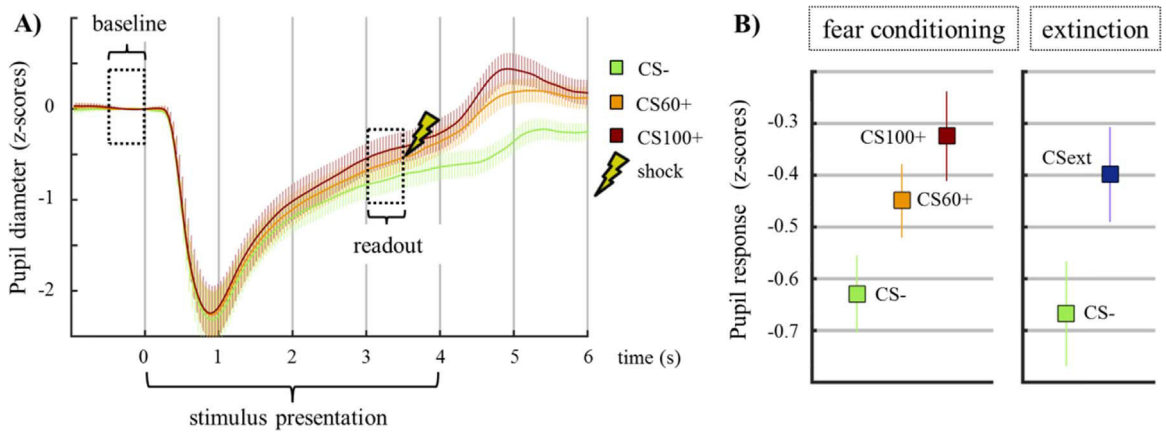


Fig. 1. A) Average change in pupil diameter in response to CS-, CS60+ and CS100+ across trials of the fear conditioning phase with 95% confidence intervals (N=28). The stimulus was presented from 0 to 4 seconds, shock administration occurred at 3.5 s in all CS100+ trials and in 60% of CS60+ trials. For assessing the pupil response per trial, pre-stimulus baseline (marked as 'baseline', average across 0.5 s pre-stimulus) was subtracted from the maximum pupil diameter between 3 and 3.5 s after stimulus onset (marked as 'readout'). **B)** Mean pupil responses during fear conditioning and extinction are depicted as boxes with standard error of the mean (± SEM) as vertical lines (N=28). The extinguished stimuli were collapsed as CSext, comprising both extinction groups (extinction of CS60+ and CS100+), as they did not differ significantly.

by Beatty and Lucero-Wagoner (2000). Another readout of interest concerned the reflexive pupil constriction at CS onset, as Reinhard et al. (2006) found significantly stronger inhibition of the initial light reflex at the onset of CS+ than of CS-. We assessed the initial light reflex by subtracting the respective baseline diameter value from the pupil diameter trough in the first second after stimulus onset.

To furthermore consider task-unrelated tonic drifts in pupil diameter, we determined the 46 pre-stimulus pupil baselines of all trials, again treating values from invalid trials as missing data.

2.4. Statistical analysis of behavioral data and pupil responses

Statistical analyses of behavioral data and pupil responses were carried out in SPSS (PASW Statistics for Windows, Version 18.0. Chicago: SPSS Inc.). Subjective ratings were analyzed by conducting a stimulus (CS-, CS60+, CS100+)×time (5 ratings) repeated measures analysis of variance (rmANOVA). To assess the effects of partial reinforcement on extinction, a separate stimulus (CS-, CSext)×time (rating 3, 4 and 5)×group (extinction of CS60+ vs. CS100+) rmANOVA was conducted.

Pupil responses were analyzed for the fear conditioning (trial 1–30) and extinction (trial 31–46) phases of the task separately. For fear conditioning, we conducted an rmANOVA on the mean pupil responses to CS-, CS60+ and CS100+, as well as post-hoc t-tests for directed comparisons of the three stimulus types. To investigate the effect of full versus partial reinforcement more closely, post-hoc t-tests between the mean pupil response magnitude to CS60+ and CS100+ were performed for the first and second half of fear conditioning separately. Moreover, dynamics of pupil responses to CS100+ were explored with a separate rmANOVA over the 10 trials of the fear conditioning phase for this stimulus.

For the extinction phase, we conducted a stimulus × group (extinction of CS60+ vs. CS100+) rmANOVA on the mean pupil responses to CS- and CSext, in a similar manner as the behavioral rmANOVA for the extinction phase. A post-hoc t-test over the mean response to CS- and CSext was performed, as well as post-hoc t-tests to compare the mean response to CS- and CSext between extinction groups. In addition we performed paired t-tests on the initial light reflexes to CS-, CS60+ and CS100+.

To investigate effects of habituation, an rmANOVA was performed over all 46 pupil responses of the entire fear learning task (irrespective of stimulus type), as well as over all 46 pre-stimulus baselines. Moreover, we examined the correlation of pre-stimulus baselines with subsequent pupil responses, as a negative correlation between these two has been reported (Eldar et al., 2013). We calculated Pearson correlations for each subject individually and, to assess if these correlations were significant on the group level, we performed a non-parametric Sign test.

To account for this anticipated correlation of pre-stimulus baselines with the following pupil responses, we furthermore calculated baseline-weighted pupil responses for an additional fMRI analysis. We linearly transformed all pupil responses and all baselines to values from 0.01 to 0.99 (to avoid zero values). We then multiplied pupil responses with the respective baseline values, resulting in 46 baseline-weighted pupil responses per subject. These were again split into the three stimulus types for further analyses.

2.5. fMRI

2.5.1. Image acquisition and preprocessing

Scanning was performed on a 3 T MR scanner (Discovery MR750, GE Healthcare, Waukesha, WI, USA) using a 32-channel head coil. For the acquisition of functional data, the whole brain was covered with 36 axial slices in AC-PC orientation (TR=2.56 s, acceleration factor of 2, FOV=22 cm, 64×64 matrix, slice thickness 3 mm and 0.4 mm gap leading to a voxel size of 3.4×3.4×3.4 mm^3). A multi-echo planar imaging sequence (MEPI) with three echo pulses was used, however, only images resulting from the second echo (TE=29 ms) were analyzed for this study.

Preprocessing was performed with SPM8 (Statistical Parametric Mapping Software, www.fil.ion.ucl.ac.uk/spm). Data were slice time corrected, realigned to the mean volume (rigid-body transformation), normalized to the Montreal Neurological Institute (MNI) EPI template, spatially smoothed (full width at half maximum Gaussian kernel of 6×6×6 mm^3), and high pass filtered at 0.004 Hz. The entire fear learning task comprised a total of 384 volumes (236 volumes corresponding to fear conditioning, 148 to extinction), of which the first four volumes (i.e. 10.24 s) were excluded from further analysis to avoid non steady-state effects. To account for signal fluctuations of non-neural origin, nuisance regressors reflecting fluctuations in cerebrospinal fluid (CSF) and white matter (WM) compartments were calculated using the CompCor method (Behzadi et al., 2007). For this purpose, realigned fMRI data were normalized to a resolution of 4×4×4 mm^3 and the first three principal components were extracted from thresholded SPM probability maps for CSF and WM compartments. These six time series, as well as six regressors for head motion (extracted during the realignment process) were incorporated into all further fMRI analyses, along with the absolute values of their first order derivatives. These nuisance regressors were complemented by one regressor including all shock onsets and one regressor for the subjective ratings, resulting in a total of 26 nuisance regressors.

2.5.2. Statistical analysis of fMRI data

For analyzing activity related to stimulus presentations, preprocessed volumes were entered into a general linear model (GLM) that contained three regressors of interest corresponding to the presentations of CS−, CS60+ and CS100+. Each stimulus presentation was modeled as an event of 3.4 seconds duration. On the group level, we contrasted activity relating to CS- and CS+ (both CS60+ and CS100+) with post-hoc t-tests in a full-factorial ANOVA (contrasts [−2 +1 +1] and [+2 −1 −1]). These analyses were performed for the entire fear learning task, as well as for the fear conditioning phase separately. Note that this stimulus contrast is affected by US administration, as stimulus durations were relatively short and convolvement with the hemodynamic response function (HRF) will result in overlap with US-associated activity.

For examining brain activity related to pupil responses across the three stimulus types, a second GLM was set up with one regressor containing all 46 stimulus presentations (irrespective of stimulus type) and its parametric modulation comprising all corresponding 46 pupil responses per subject. We refer to this as 'correlates of trial-by-trial pupil responses between stimulus types', because this analysis contains actual differences in response magnitude among the three stimuli. On the group level, we identified brain activity which correlated positively and negatively with this parametric modulation in a simple t-test (contrasts [+1] and [−1]).

In order to determine the neural correlates of pupil responses within each stimulus type separately, the responses were entered into a third GLM as separate parametric modulations to the corresponding three stimulus regressors (CS-, CS60+ and CS100+). We later refer to this as 'correlates of trial-by-trial pupil responses within stimulus types', as it identifies activation associated with pupil response magnitude within consecutive trials of each condition separately (for example activation related to pupil responses to CS100+ trials). In a full-factorial ANOVA on the group level, brain activity positively and negatively correlated with the three parametric modulations was identified with post-hoc t-tests (contrasts [+1 +1 +1] and [−1 −1 −1]). Furthermore, post-hoc t-tests to compare statistical maps were performed (contrasting CS- versus CS+, [-2 +1 +1] and [+2 −1 −1]). The same procedure was performed for the baseline-weighted pupil responses in a fourth GLM. This analysis addressed the effect of slow tonic changes in baseline pupil diameter on the magnitude of pupil responses. These within-stimulus analyses of pupil responses and baseline-weighted pupil responses were performed for the entire fear learning task, as well as for the fear conditioning phase separately. For all GLMs examining the correlates of pupil responses, invalid trials were removed from the stimulus regressors and were entered into the GLMs as additional nuisance regressors (one regressor per stimulus type with no parametric modulation).

Finally, to compare tonic changes in pupil diameter to general changes in brain activity, a GLM with one regressor per stimulus type (CS-, CS60+ and CS100+) and their respective linear time modulation was performed over trials of the fear conditioning phase. On the group level, a post-hoc t-test on a full-factorial ANOVA was employed to detect activity that was linearly declining throughout the fear conditioning phase (contrast [−1 −1 −1]).

For all fMRI analyses, clusters were sampled at p=.001 (uncorrected) and significance was defined as cluster p-values < .05 after whole-brain family-wise error (FWE) correction.

3. Results

3.1. Subjective ratings

Subjective ratings averaged over all 34 subjects are displayed in Fig. 2A. The two-way rmANOVA of all 5 ratings yielded significant main effects of stimulus type ($F_{(2,66)}$=264.25, p < .001), of time ($F_{(4,132)}$=34.79, p < .001) and a significant stimulus×time interaction ($F_{(8,264)}$=32.88, p < .001). The three-way rmANOVA of ratings during extinction (ratings 3–5) yielded significant main effects of stimulus type ($F_{(1,32)}$=218.31, p < .001), time ($F_{(2,64)}$=192.30, p < .001) and group ($F_{(1,32)}$=7.42, p < .05), as well as significant interactions of stimulus×group ($F_{(1,32)}$=6.31, p < .05), time×group ($F_{(2,64)}$=14.22, p < .001), stimulus×time ($F_{(2,64)}$=188.70, p < .001) and a three-way interaction of stimulus×time×group ($F_{(2,64)}$=13.72, p < .001). However, as can be seen in Fig. 2A, these differences were not in the expected direction of the PREE (ratings 3–5 reflect group values of N=19 for CS60+ and N=15 for CS100+; ratings of CS- are collapsed across both groups).

3.2. Phasic pupil responses

For the analysis of the entire fear learning task, a total of 26 subjects fulfilled the inclusion criteria. For separate analyses of the fear conditioning and extinction phases, 28 subjects each could be included (the rest having over 20% discarded trials; partly differing subjects for the fear conditioning and extinction phases).

The rmANOVA of mean pupil responses during fear conditioning yielded a significant effect of stimulus ($F_{(2,54)}$=8.75, p < .005). Post-hoc t-tests revealed stronger pupil responses to both CS+ than to CS-

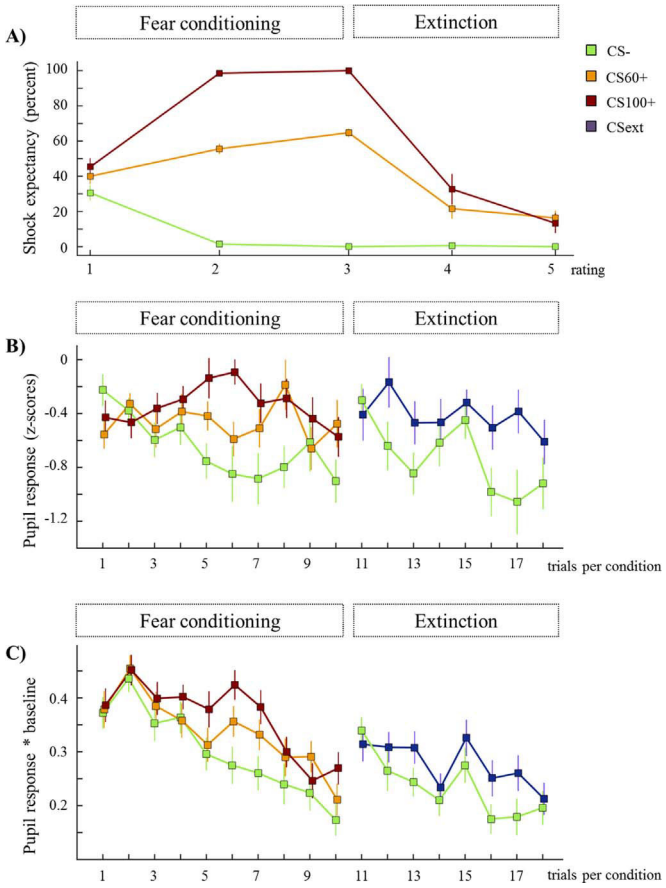


Fig. 2. A) Average subjective ratings of shock probability for each stimulus type (CS-, CS60+ and CS100+) with respective SEMs. Ratings occurred before fear conditioning, halfway through fear conditioning (after 5 trials of each stimulus type), before extinction (after 10 trials of each stimulus type), halfway through extinction (after 4 additional trials per stimulus type) as well as after extinction. Ratings 1 and 2 comprise all subjects (N=34); ratings 3, 4 and 5 reflect scores of all 34 subjects for CS- and group scores of N=19 for CS60+ and N=15 for CS100+. **B)** Trial-wise pupil responses to CS-, CS60+ and CS100+ with SEMs during fear conditioning and extinction phases (average across subjects, N=26). **C)** Trial-wise pupil responses, weighted by respective pre-stimulus baselines to CS-, CS60+ and CS100+ with SEMs (average across subjects, N=26). In B) and C), the extinguished CS60+ and CS100+ were collapsed as CSext, because there were no significant differences in response to those stimuli between groups.

(CS60+ > CS-: $t_{(27)}$=2.51, p < .01; CS100+ > CS-: $t_{(27)}$=4.15, p < .001; also see Fig. 1B). Pupil responses to CS100+ and to CS60+ diverged during the fear conditioning phase (all trials: $t_{(27)}$=1.64, p=.056; first half: $t_{(27)}$=1.23, p=.22, second half: $t_{(27)}$=1.80, p < .05). Pupil responses to CS100+ showed a significant quadratic within-subject contrast for these trials ($F_{(1,13)}$=8.37, p < .05), reflecting an initial increase and a later decrease in response magnitude to this stimulus during fear conditioning (Fig. 2B).

For extinction, the rmANOVA of mean pupil responses to CS- and CSext with the factor group yielded a significant main effect of stimulus type ($F_{(1,26)}$=12,63, p < .005), but no significant effect of group ($F_{(1, 23)}$=0.49, p=.491) and only a trend for a stimulus×group interaction ($F_{(1, 26)}$=2.94, p=.098). Post-hoc t-tests revealed significantly higher responses to CSext than to CS- ($t_{(27)}$=3.32, p < .005), but no significant differences between groups in the response to CSext ($t_{(26)=}$0.14, p=.444), which is why CS60+ and CS100+ are collapsed as CSext in Fig. 2B and C. There was a trend for stronger pupil responses to CS- in the group in which CS60+ was extinguished ($t_{(26)}$=1.41, p=.084). The initial light reflex to CS-, CS60+ and CS100+ did not yield any significant differences between stimuli (p > .05, d < .20). Averaged trial-by-trial pupil responses during fear conditioning and extinction phases are illustrated in Fig. 2B (mean across 26 subjects).

3.3. Tonic changes in pupil diameter

Although there was no significant effect of time for the 46 pupil responses ($F_{(45,135)}$=1.32, p=.115), there was a significant effect of time for the 46 pre-stimulus baseline values ($F_{(45, 135)}$= 1.96, p < .005, see Supplementary Fig. S1). Pupil responses to the stimuli were negatively correlated with the size of pre-stimulus baseline diameter (Sign test, p < .001, Pearson's R averaged across 26 subjects=−.33). To control for this negative correlation, we multiplied pupil responses by their respective pre-stimulus baselines (both linearly transformed to values between 0.01 and 0.99), and the resulting baseline-weighted pupil responses (see Fig. 2C) were used for further fMRI analysis.

3.4. fMRI results

For fear conditioning, the contrast CS- < [CS60+ and CS100+] revealed significant clusters of activation in left and right insula, extending bilaterally into rolandic operculum and supramarginal gyrus, a large cluster extending from dACC to the supplementary motor area (SMA) and into the paracentral lobule, a cluster covering left and right thalamus and a bilateral cluster in calcarine gyrus and lingual gyrus (see Fig. 3 and Supplementary Table 1). To disentangle contributions of different regions within the bilateral clusters around insula, the collection threshold was increased to p=.0001, which revealed a separate cluster comprising right amygdala, caudate and hippocampus (p=.014, t = 4.68, cluster size k=117, peak at [20, −2, 16]).

The reverse contrast of CS- > [CS60+ and CS100+] revealed activity in left middle temporal gyrus, as well as two clusters in the medial prefrontal cortex (mPFC): one in vmPFC comprising orbitofrontal cortex and gyrus rectus, and another in superior mPFC (see Fig. 3 and Supplementary Table 2). Note that this stimulus contrast is affected by US-associated activity. It is presented for illustration purposes and for visual comparison with the statistical maps of pupil related activity (see Fig. 4), as well as with previous work.

3.5. Neural correlates of pupil responses

3.5.1. Correlates of trial-by-trial pupil responses between stimulus types

The parametric modulation with trial-wise pupil response magnitudes during the entire fear learning task revealed significant activity in dACC extending to SMA, bilateral insula, thalamus, rolandic operculum extending to supramarginal gyri and a cluster in left precentral gyrus. The same parametric modulation revealed deactivation in bilateral occipital cortex and vmPFC (see Fig. 4A and Tables 1 and 2). Note that this contrast contains differences between the three stimulus types (e.g., generally larger pupil responses to CS+ than to CS-, as shown in Fig. 1B).

3.5.2. Correlates of trial-by-trial pupil responses within stimulus types

Therefore, we next entered trial-wise pupil responses as parametric modulations to each of the three stimulus regressors separately, such that mean differences in response magnitude between the stimulus types were removed. For the entire fear learning task, these three parametric modulations (contrast [+1 +1 +1]) revealed activation in dACC (two clusters, one of which extended to SMA), in right thalamus, as well as activity in right rolandic operculum and supramarginal gyrus (see Fig. 4B and Table 3). A cluster in right insula showed a trend for significance (p=.087, t=4.95, cluster size k=107, peak at [36, 18, −8]). Direct overlap between this statistical map and the stimulus contrast (CS+ > CS-) throughout the entire fear learning task was observed in dACC and right supramarginal gyrus, see Supplementary Fig. S2 for a visual comparison. The same parametric modulation yielded deactivation in bilateral middle occipital gyrus (contrast [−1 −1 −1], see Fig. 4B and Table 4). The analysis for the fear conditioning phase alone still revealed activation in dACC (contrast [+1 +1 +1], see Supplementary Fig. S3A, p=.011, t=4.10, k=176, peak at [2, 20, 38]). Contrasting the neural correlates of the pupil responses to CS- and CS+ in post-hoc t-

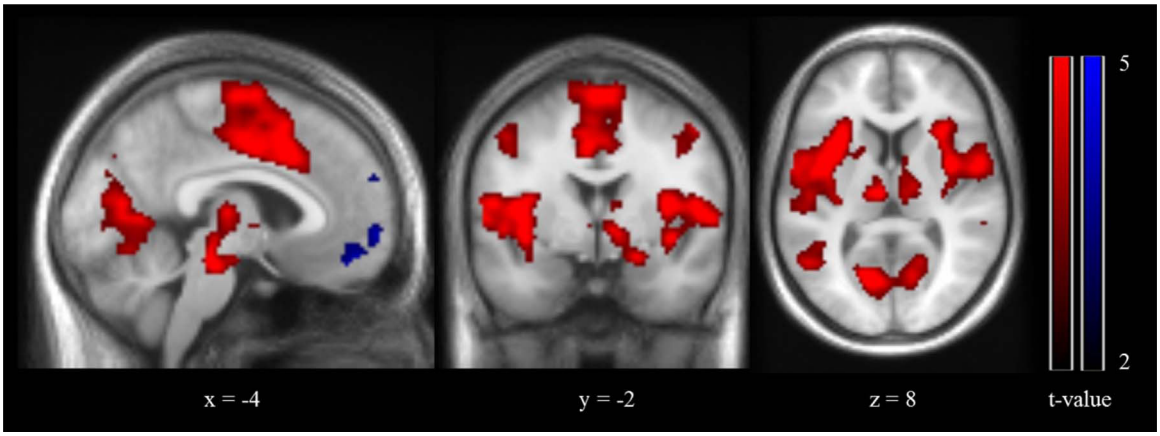


Fig. 3. Activity associated with the stimulus contrast CS- < [CS60+ and CS100+] during the fear conditioning phase (red), and the reverse contrast in blue. Post-hoc t-contrasts were derived from a full-factorial ANOVA with the factor stimulus with the levels CS-, CS60+ and CS100+ (contrasts: [−2 +1 +1] and [+2 −1 −1], N=34). Coordinates [mm] refer to MNI space.

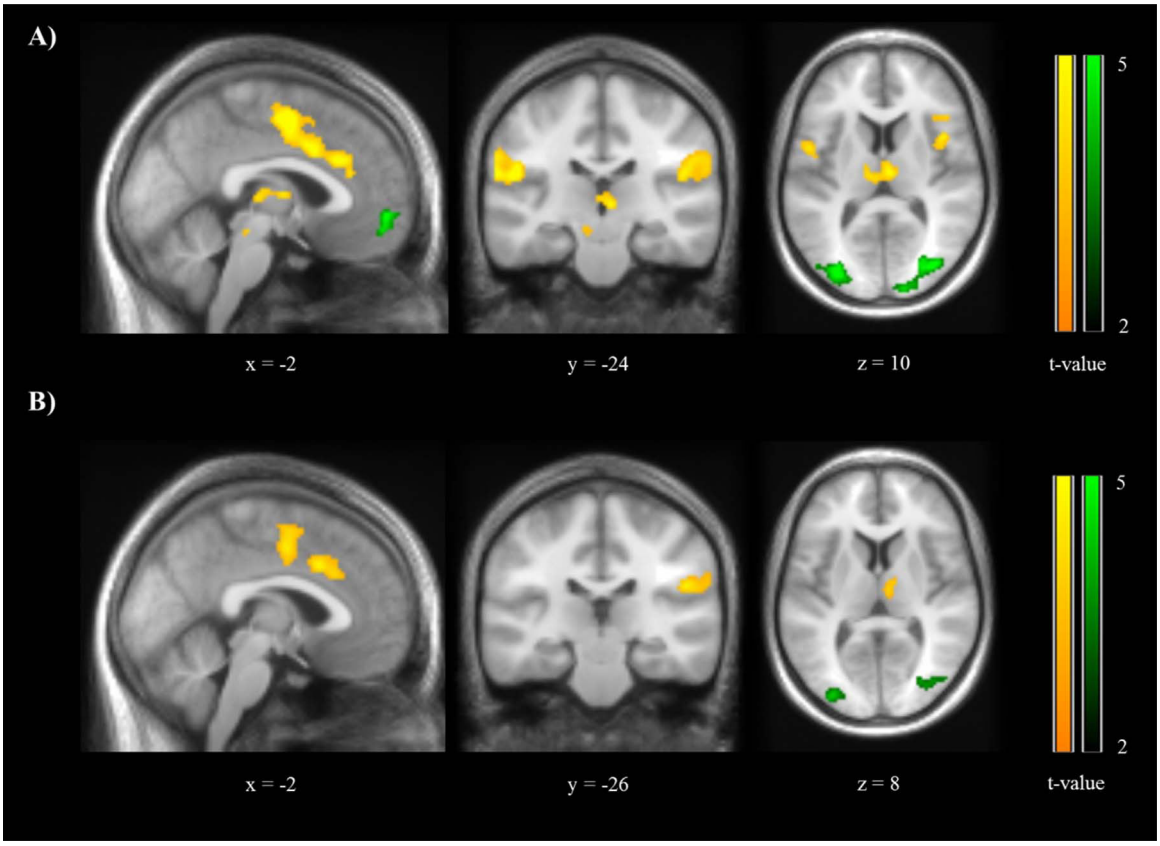


Fig. 4. A) Activity positively (yellow) and negatively (green) correlated with the magnitude of trial-by-trial pupil responses containing differences between stimulus types (parametric modulation to all stimulus presentations of CS-, CS60+ and CS100+ in one regressor; contrasts [+1] and [−1], N=26). **B)** Activity positively (yellow) and negatively (green) correlated with the magnitude of trial-wise pupil responses within the three stimulus types. Contrasts were derived from a full-factorial ANOVA of the parametric modulations to CS-, CS60+ and CS100+ in separate regressors; (contrasts [+1 +1 +1] and [−1 −1 −1], N=26). Coordinates [mm] refer to MNI space.

tests (e.g., contrast [−2 +1 +1] and [+2 −1 −1]) revealed no significant differences.

3.5.3. Correlates of trial-by-trial baseline-weighted pupil responses and time modulations

The parametric modulation with baseline-weighted pupil responses (showing decreasing slopes for all stimuli, see Fig. 2C) to the three corresponding stimulus regressors separately revealed no significant clusters for the entire fear learning task. However, for the fear conditioning phase, this parametric modulation yielded activity in right thalamus (see Supplementary Fig. S3B and S3C, p=.006, t=4.91, cluster size k=200, peak at [14, 6, 0]).

The same thalamic cluster was negatively correlated with a linear time modulation over trials of the fear conditioning phase (contrast [−1], reflecting decreasing activity during the fear conditioning phase). This declining time modulation further revealed activity in bilateral amygdala and hippocampus, among other regions (see Supplementary Fig. S3C and Supplementary Table S3).

4. Discussion

4.1. Phasic pupil responses during fear conditioning and extinction

Our results confirm previous findings that pupil dilations can serve as a physiological readout of the conditioned response (Reinhard et al., 2006; Visser et al., 2013). We found that CS+ elicited stronger pupil responses than CS- during the fear conditioning as well as during the extinction phase, largely in line with subjective ratings of shock probability. The stimulus contrast CS+ > CS- was, as expected, associated with an activation pattern resembling the fear network as reported by Etkin and Wager (2007) and Fullana et al. (2015), with contributions of bilateral insula, dACC (extending to SMA), supramarginal gyrus, thalamus and right amygdala.

The activity pattern associated with pupil response magnitude across the entire fear learning task showed substantial overlap with this network, revealing a generally stronger association of fear network activity with pupil responses to CS+ than to CS-. This finding was expected, as the analysis comprising between-stimulus differences in pupil response magnitude should resemble the CS+ > CS- contrast. We controlled for this by analyzing activity related to trial-wise pupil responses for each stimulus type separately. This analysis also revealed activity in parts of the fear network, with significant clusters of activity in dACC, right thalamus and supramarginal gyrus. This activation pattern was found across stimulus types, indicating that variability in pupil response magnitude is generally associated with activity in dACC in a fear learning context. The same analysis for the conditioning phase alone still revealed dACC activity, so we can exclude that these results were merely caused by differences in responding between the fear conditioning and extinction phases of the task.

4.1.1. Fear conditioning

Phasic pupil responses during fear conditioning were larger to CS+ than to CS- and were – irrespective of stimulus type – most robustly associated with dACC activity. Spontaneous pupil dilations during rest have already been linked to ACC activity in the macaque monkey (Aston-Jones and Cohen, 2005; Ebitz and Platt, 2015; Joshi et al., 2016) and also to activity in dACC in humans (Murphy et al., 2014; Schneider et al., 2016). Moreover, other autonomous responses to fear stimuli have previously been related to dACC activity: the difference in SCR to CS+ and CS- in fear conditioning (Milad et al., 2007), as well as

Table 1
Activity positively correlated with pupil response magnitude during the entire fear learning task, containing differences between stimulus types. The contrast was derived from a GLM containing one parametric modulation to all stimulus onsets of CS-, CS60+ and CS100+ (contrast [+1]; N=26).

Region name	Cluster		Voxel			
	P_{corr}	k	T_{peak}	MNI_{peak}		
				x	y	z
Middle cingulate cortex (L+R) Supplementary motor area (L+R) Anterior cingulate cortex (L+R)	< .001*	1305	8.10	−4	-6	50
Insula (R) Inferior frontal gyrus (R) Inferior frontal gyrus, pars opercularis & triangularis (R)	< .001*	595	6.93	34	6	−12
Insula (L) Superior temporal gyrus (L)	.004*	184	6.90	−36	−4	−10
Supramarginal gyrus (L) Postcentral gyrus (L) Superior temporal gyrus (L) Rolandic operculum (L)	< .001	491	6.28	−50	−22	20
Supramarginal gyrus (R) Rolandic operculum (R)	< .001	368	5.78	48	−26	18
Thalamus (R+L)	< .001	583	5.64	8	−4	2
Rolandic operculum (L) Precentral gyrus (L) Superior temporal pole (L) Inferior frontal gyrus, pars opercularis (L)	.020	131	5.24	−50	0	14
Precentral gyrus (L) Superior frontal gyrus (L) Paracentral lobule (L)	.019	133	5.16	−18	−4	72

Note: Table 1 refers to Fig. 4A (yellow clusters). Only brain regions contributing over 5% to respective clusters are listed. R=right; L=left; P_{corr} stands for whole-brain corrected cluster p-values, k for the cluster size, t_{peak} for the t-value of the peak-voxel, [x y z] coordinates are in MNI-space. Asterisks mark clusters which are significant after voxel-wise inference.

Table 2
Activity negatively correlated with pupil response magnitude during the entire fear learning task, containing differences between stimulus types. The contrast was derived from a GLM containing one parametric modulation to all stimulus onsets of CS-, CS60+ and CS100+ (contrast [−1]; N=26).

Region name	Cluster		Voxel			
	P_{corr}	k	T_{peak}	MNI_{peak}		
				x	y	z
Middle occipital gyrus (L) Inferior occipital cortex (L) Fusiform gyrus (L)	< .001*	1058	6.98	−24	−80	18
Middle occipital gyrus (R) Inferior occipital cortex (L) Superior occipital gyrus (R) Cuneus (R)	< .001	821	5.70	36	−80	10
Medial orbitofrontal cortex (L+R)	.029	121	5.22	−2	52	−6

Note: Table 2 refers to Fig. 4A (green clusters). Only brain regions contributing over 5% to respective clusters are listed. R=right; L=left; P_{corr} stands for whole-brain corrected cluster p-values, k for the cluster size, t_{peak} for the t-value of the peak-voxel, [x y z] coordinates are in MNI-space. Asterisks mark clusters which are significant after voxel-wise inference.

Table 3
Activity positively correlated with pupil response magnitude within stimulus types during the entire fear learning task. The contrast was derived from a full-factorial ANOVA of parametric modulations to the corresponding stimulus types CS-, CS60+ and CS100+ separately (contrast [+1 +1 +1], N=26).

Region name	Cluster		Voxel			
	P_{corr}	k	T_{peak}	MNI_{peak}		
				x	y	z
Middle cingulate cortex (L+R) Anterior cingulate cortex (L+R) Thalamus (R)	< .001*	502	5.24	−6	12	36
	.024	150	4.77	8	−6	2
Supramarginal gyrus (R) Rolandic operculum (R)	.002	245	4.69	50	−24	24
Middle cingulate cortex (L+R) Supplementary motor area (L+R)	< .001	388	4.53	−2	−6	46

Note: Table 3 refers to Fig. 4B (yellow clusters). Only brain regions contributing over 5% to respective clusters are listed. R=right; L=left; P_{corr} stands for whole-brain corrected cluster p-values, k for the cluster size, t_{peak} for the t-value of the peak-voxel, [x y z] coordinates are in MNI-space. Asterisks mark clusters which are significant after voxel-wise inference.

Table 4
Activity negatively correlated with pupil response magnitude within stimulus types during the entire fear learning task. The contrast was derived from a full-factorial ANOVA of parametric modulations to the corresponding stimulus types CS-, CS60+ and CS100+ separately (contrast [−1 −1 −1], N=26).

Region name	Cluster		Voxel			
	P_{corr}	k	T_{peak}	MNI_{peak}		
				x	y	z
Middle occipital gyrus (R) Inferior occipital cortex (R)	.001	261	4.74	44	−82	−6
Middle occipital gyrus (L)	.013	173	4.73	−32	−88	10

Note: Table 4 refers to Fig. 4B (green clusters). Only brain regions contributing over 5% to respective clusters are listed. R=right; L=left; P_{corr} stands for whole-brain corrected cluster p-values, k for the cluster size, t_{peak} for the t-value of the peak-voxel, [x y z] coordinates are in MNI-space. Asterisks mark clusters which are significant after voxel-wise inference.

the magnitude of SCR in the anticipation of pain (Seifert et al., 2013) were found to positively correlate with dACC activity. Prior evidence for a relationship between task-related pupil dilations and dACC activity was also provided by Critchley et al. (2005), who found activity in rostral ACC and dACC to correlate with the magnitude of pupil responses in a numerical stroop task. However, in other studies, pupil responses have been associated with different circuity as well: they have been associated with activation of superior temporal gyrus in a speech comprehension task (Zekveld et al., 2014), with activity in middle frontal gyrus in a digit sorting working memory task (Siegle et al., 2003) and indirectly with the dorsal attention network (together with LC, superior colliculus and right thalamus) in an attention task (Alnaes et al., 2014). Our study provides evidence for a task-specific association of pupil responses with dACC in a fear learning task.

Recent findings from the macaque brain revealed that ACC activity can follow pupil dilations, but can also temporally precede them (Joshi et al., 2016), hence both directions could account for this association. Although our data cannot provide evidence for a causal influence of dACC on pupil responses, we propose that at least part of this association could be explained by threat appraisal, which has been related to dACC activity in previous work (Etkin et al., 2011; Maier et al., 2012; Mechias et al., 2010). In line with this, we found that pupil dilations in response to CS+ appear most robustly seconds after

stimulus onset, just before US administration. This rather late onset argues against a reflexive, non-conscious fear response to CS+, and supports the notion of conscious US expectancy or threat appraisal influencing pupil diameter. The discrimination of responses to the fully and partially reinforced CS+ during fear conditioning provide further support for the proposed relationship between threat appraisal and pupil dilation.

However, responses to CS100+ declined significantly during the second half of fear conditioning, unlike the responses to CS60+ and unlike subjective ratings of shock expectancy. We therefore suggest that pupil response magnitude in our task is also influenced by uncertainty or unpredictability, which have previously been found to correlate with pupil diameter (Lavin et al., 2014; Morriss et al., 2015; Nassar et al., 2012; Preuschoff et al., 2011). In our task, subjects may become increasingly confident on the outcome of CS100+ trials during the fear conditioning phase (resulting in less uncertainty), while they need to constantly update their expectations for CS60+ trials. Therefore, dACC activity, related to the magnitude of trial-by-trial pupil responses, may be interpreted as representing threat appraisal but also uncertainty.

Given the general association of pupil response magnitude (irrespective of stimulus type) with dACC activity and with parts of the fear network, pupil responses during fear learning may also more generally reflect saliency. Both the fear network, but also a more general 'salience network' (sometimes referred to as the 'cingulo-opercular network') comprise dACC, insula and thalamus (Coste and Kleinschmidt, 2016; Sadaghiani and D'Esposito, 2015). The salience network has been proposed to respond, independently of task requirements, to personally salient stimuli (Seeley et al., 2007) and to relate to states of alertness (Coste and Kleinschmidt, 2016; Sadaghiani and D'Esposito, 2015). Hence, pupil responses may track the general saliency of stimuli across different phases of fear learning on a trial-by trial basis.

4.1.2. Extinction

During extinction, pupil responses remained higher for CSext than for CS-. This suggests that autonomous arousal is maintained towards the previously reinforced stimulus, and that pupil responses may be a more sensitive readout than other measures like SCR. The increased pupil responses towards CSext may reflect elevated shock expectancy, although pupil responses during extinction did not match the subjective ratings exactly. Ratings differed significantly for the fully and partially reinforced stimulus between groups whereas pupil responses did not, although this may be due to less precision of the latter. Another explanation would be that increased uncertainty contributes to pupil responses to both CSext, irrespective of their previous reinforcement rate, since neither was followed by the anticipated shocks anymore. In line with this notion, both pupil responses to CS- and baseline diameter appeared to increase in early extinction trials. Uncertainty may be particularly increased in early extinction, when subjects notice a change in reinforcement rules and may expect a reversal of CS-US contingencies. Contrary to our expectations, we did not observe a PREE (partial reinforcement extinction effect) as described by Grady et al. (2016), neither in pupil responses nor in subjective ratings.

4.2. Tonic pupil diameter

While phasic pupil responses to the stimuli did not show habituation over time, tonic pupil diameter (as measured by the pre-stimulus baseline) decreased significantly during the entire fear learning task. Furthermore, pre-stimulus baseline pupil diameter correlated negatively with the magnitude of phasic pupil responses to the stimuli. This inverse relationship has been described before (Eldar et al., 2013; Joshi et al., 2016) and may be due to a larger margin for dilation given a small initial pupil size. We controlled for this association by weighting (i.e. multiplying) pupil responses with their respective baseline diameter. These baseline-weighted pupil responses, corrected for the tonic pupil diameter, showed a declining slope and were correlated with thalamic activity during fear conditioning. A linearly decreasing time modulation to trials of the fear conditioning phase revealed the same thalamic cluster, which suggests that both tonic pupil diameter and thalamic activity decreased in the conditioning phase.

Even though thalamus activity was associated with phasic pupil responses (together with dACC), we propose that both the decrease in tonic pupil diameter and the associated decreasing thalamic activity may reflect habituation or a decrease in general alertness. Anatomical considerations support this notion: the thalamus is located within the ascending reticular activating system (ARAS), which originates in the brainstem and has extensive wakefulness-promoting projections throughout the brain (Aston-Jones and Cohen, 2005). A decrease in thalamic activity has previously been demonstrated to occur during the transition from wakefulness to sleep (Kaufmann et al., 2006). Moreover, spontaneous tonic fluctuations in pupil diameter have previously been used as an indicator of wakefulness and have been referred to as 'pupillary unrest' (Lowenstein et al., 1963; Wilhelm et al., 1998). Such changes in tonic pupil diameter during rest have been linked to the thalamus before (Schneider et al., 2016), with increasing thalamic activity at the onset of spontaneous pupil dilations. The LC can possibly serve as a mediating structure with the potential to cause both changes in thalamic activity - via dense norepinephrinergic projections (Samuels and Szabadi, 2008) - and wakefulness-related changes in pupil size (Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010; Joshi et al., 2016; Murphy et al., 2014).

To summarize, tonic and phasic components of the pupil diameter may contain complementary information. The tonic component can be used for assessing habituation or controlling for wakefulness and states of alertness, which is a major advantage of pupillometry during simultaneous fMRI. Meanwhile, phasic pupil responses did not show significant habituation throughout the task, and continued to discriminate between CS- and CSext even in late extinction trials. This constitutes another advantage in comparison to other physiological measures of arousal, like SCR or startle responses. If the tonic pupil size is thought to differently affect pupil responses to different stimuli or during different phases of a task, weighting pupil responses by their respective baseline values could control for this. Furthermore, the neural correlates of tonic pupil diameter and phasic pupil responses can be partly disentangled. It therefore seems useful to take both components into account when interpreting pupil dilations as a readout of cognitive processes.

4.3. Limitations

Future research could further disentangle sympathetic versus parasympathetic influences on pupil responses to CS+ and CS-. We assume that pupil dilations in fear conditioning are largely driven by an increase in sympathetic activation (and possibly inhibition of parasympathetic input). However, from our data we cannot conclude if the smaller pupil responses to CS- are due to the absence of sympathetic arousal, or to active fear inhibition and autonomous downregulation in response to CS-. Pollak et al. (2010) found pupillary constriction during learned safety, however in our data such effects would have been masked by the strong initial pupillary light reflex. We found pupil responses to be negatively associated with activity in vmPFC, a region that has previously been associated with safety signaling and extinction (Milad and Quirk, 2012). A lesion study in humans indicated an influence of ventromedial prefrontal cortex on pupil dilation during reward processing (Manohar and Husain, 2016), supporting the notion that vmPFC activity may in fact be responsible for active pupillary constriction during safety signaling. Yet in our analyses, the cluster in vmPFC did not reach significance when examining the correlates of pupil constrictions for each stimulus separately. This indicates that the relationship between vmPFC and pupil responses in our data may mostly reflect the CS- to CS+ difference. In future work, active

parasympathetic pupil constriction to CS- could be detected by using stimuli which are isoluminant to the background and do not cause an initial light reflex.

We did not find significant differences in the initial light reflex to the CS+ and CS-, even though this had previously been reported in fear conditioning (Reinhard et al., 2006) and for neutral versus emotionally arousing scenes (Bradley et al., 2008). Both studies have been conducted outside the MR environment, which might offer a better context for detecting potentially small differences in pupil dilation. Another limitation of this study is the lack of SCR recordings as a commonly used readout of fear conditioning. A comparison of pupil responses and SCR could relate pupillometry to other readouts of conditioned fear, although the MR environment may not be optimal to study these differences. Moreover, displaying a fixation cross in the middle of CS stimuli may reduce artifacts caused by gaze shift.

Due to the low temporal resolution of fMRI, it is difficult to determine the exact temporal alignment – and hence causality – between activity in higher-order cortical areas, arousal-regulating brainstem nuclei and pupil dilations. Beyond that, our whole-brain coverage fMRI recordings were not suitable to detect the proposed association between brain stem activity and pupil dilations. Despite many previous associations of pupil diameter and LC activity (Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010; Joshi et al., 2016; Murphy et al., 2014), we find no evidence for LC-driven pupil dilations in our fear learning task. Alternative accounts have proposed the superior colliculus (Wang and Munoz, 2015) or the paragigantocellularis nucleus (Joshi et al., 2016; Nieuwenhuis et al., 2011) to be other suitable structures for relaying cortical signals to both the pupil as well as to LC, which may be investigated by targeting brain stem regions more closely.

4.4. Temporal dissection of the conditioned response

It is worth noting that we did not find evidence for a direct association between pupil dilation responses and amygdala activity; we only observed amygdala activity in our analysis with linearly declining time modulations to the three stimuli during the fear conditioning phase. The amygdala has been shown to quickly habituate during fear conditioning (Büchel et al., 1998; Lindner et al., 2015) and may be primarily involved in early phases of conditioning. Akin to amygdala activity, both SCR and fear potentiated startle also show strong habituation (Bradley et al., 1993; Pineles et al., 2009). This supports the notion that these psychophysiological conditioned responses may partly be driven by the amygdala. Further support for such a relationship comes from Van Well et al. (2012), who reported that subjects with strong startle potentiation towards CS+ revealed stronger amygdala activity to CS+ than subjects who displayed reduced startle potentiation.

Regarding SCR, such quick habituation seems to apply specifically to the so-called 'first interval response' at stimulus onset, as opposed to later SCR (Pineles et al., 2009). Cheng et al. (2007) reported that early rather than late components of SCR are associated with amygdala activity. However, the early SCR component has also been proposed to partly reflect an orienting response (Öhman, 1971; Pineles et al., 2009). Although no work has yet specifically addressed the neural correlates of late SCR components during human fear learning, a few studies have examined the SCR at CS+ offset in non-reinforced trials. These studies revealed activity in dACC, among other regions (Linnman et al., 2011; Spoormaker et al., 2011; for an overview of various temporal components of the SCR see Spoormaker et al., 2012). Such findings suggest that early and late components of the conditioned response may be associated with different neuronal circuity, with a shift from immediate amygdala activity representing automated threat detection to activity in dACC, which may reflect more conscious threat appraisal (Öhman, 2005).

Pupillometry may capture both the early components of the conditioned response (e.g., with inhibition of the initial light reflex as demonstrated by Reinhard et al., 2006) and late components of the conditioned response, comparable to early and late SCR. In our study, slow pupil dilations during fear learning showed large differences between CS- and CS+, and trial-by-trial magnitudes were robustly correlated with dACC activity and key nodes of the salience or fear network. However, we did not detect significant inhibition of the light reflex at CS+ onset. This might be due to the MR environment, but light probes administered later during CS presentation might also yield stronger conditioned responses, similar to startle probes: Grillon et al. (1993) demonstrated that the fear potentiated startle reflex increases with temporal proximity to US onset.

5. Conclusions

Pupillometry offers a non-invasive and sensitive measure to track individual fear learning on a trial-by-trial level. Pupil dilations in response to CS+ and CS- showed the largest differences in close proximity to the expected US onset, and this difference was detectable until the end of the extinction phase. Phasic pupil responses were associated with activity in parts of the salience network, in particular with dACC. Our data suggest an influence of different cognitive-affective processes like threat appraisal and uncertainty on phasic pupil responses, and effects of habituation and general alertness on tonic pupil size. Pupillometry offers a promising method to examine and model conditioned responses and to approximate activity in the salience network.

Acknowledgements

VS acknowledges financial support by the Bavarian Academy of Sciences and Humanities. The authors would like to thank Ines Eidner and Brice Fernandez for their technical support.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuroimage.2016.11.072.

References

Alnaes, D., Sneve, M.H., Espeseth, T., Endestad, T., van de Pavert, S.H., Laeng, B., 2014. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. J. Vis. 14 (4). http://dx.doi.org/10.1167/14.4.1.

Aston-Jones, G., Cohen, J.D., 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu. Rev. Neurosci. 28, 403–450. http://dx.doi.org/10.1146/annurev.neuro.28.061604.135709.

Bach, D.R., Flandin, G., Friston, K.J., Dolan, R.J., 2009. Time-series analysis for rapid event-related skin conductance responses. J. Neurosci. Methods 184 (2), 224–234. http://dx.doi.org/10.1016/j.jneumeth.2009.08.005.

Beatty, J., Lucero-Wagoner, B., 2000. The pupillary system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.), Handbook of Psychophysiology2nd ed.. Cambridge University Press, New York, NY, US, 142–162.

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (compcor) for bold and perfusion based fmri. Neuroimage 37 (1), 90–101. http://dx.doi.org/10.1016/j.neuroimage.2007.04.042.

Benedek, M., Kaernbach, C., 2010. A continuous measure of phasic electrodermal activity. J. Neurosci. Methods 190 (1), 80–91. http://dx.doi.org/10.1016/j.jneumeth.2010.04.028.

Bitsios, P., Szabadi, E., Bradshaw, C.M., 1996. The inhibition of the pupillary light reflex by the threat of an electric shock: a potential laboratory model of human anxiety. J. Psychopharmacol. 10 (4), 279–287. http://dx.doi.org/10.1177/026988119601000404.

Bitsios, P., Szabadi, E., Bradshaw, C.M., 2004. The fear-inhibited light reflex: importance of the anticipation of an aversive event. Int. J. Psychophysiol. 52 (1), 87–95. http://dx.doi.org/10.1016/j.ijpsycho.2003.12.006.

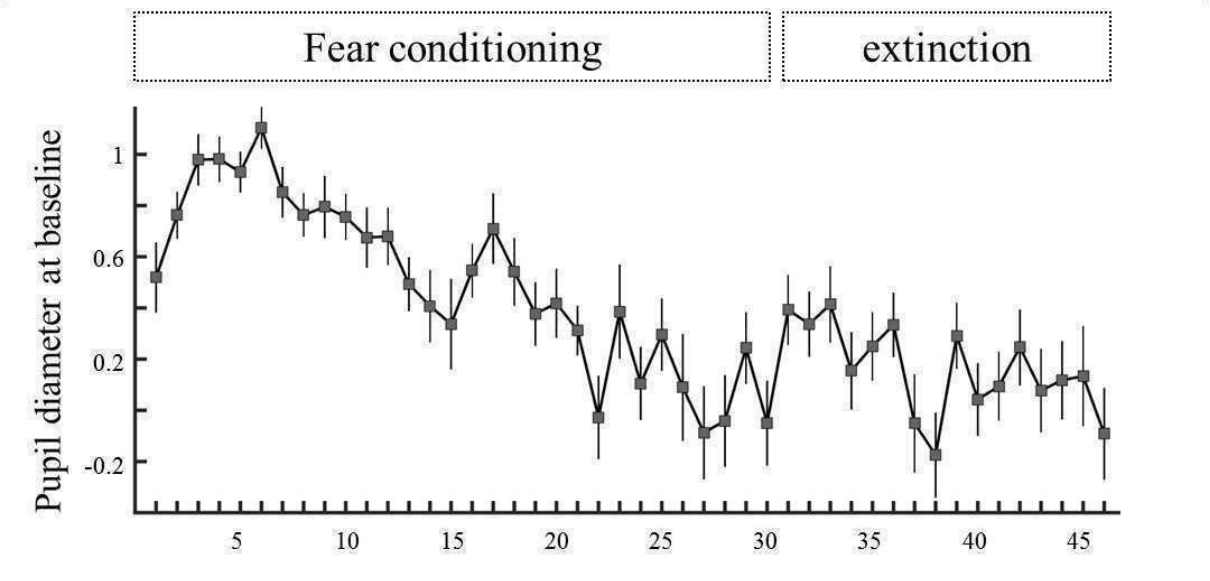
Blechert, J., Michael, T., Vriends, N., Margraf, J., Wilhelm, F.H., 2007. Fear conditioning in posttraumatic stress disorder: evidence for delayed extinction of autonomic, experiential, and behavioural responses. Behav. Res Ther. 45 (9), 2019–2033. http://dx.doi.org/10.1016/j.brat.2007.02.012.

Bradley, M.M., Lang, P.J., Cuthbert, B.N., 1993. Emotion, novelty, and the startle reflex: habituation in humans. Behav. Neurosci. 107 (6), 970–980. http://dx.doi.org/10.1037/0735-7044.107.6.970.
Bradley, M.M., Miccoli, L., Escrig, M.A., Lang, P.J., 2008. The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45 (4), 602–607. http://dx.doi.org/10.1111/j.1469-8986.2008.00654.x.
Büchel, C., Morris, J., Dolan, R.J., Friston, K.J., 1998. Brain systems mediating aversive conditioning: an event-related fmri study. Neuron 20 (5), 947–957. http://dx.doi.org/10.1016/S0896-6273(00)80476-6.
Cheng, D.T., Richards, J., Helmstetter, F.J., 2007. Activity in the human amygdala corresponds to early, rather than late period autonomic responses to a signal for shock. Learn. Mem. 14 (7), 485–490. http://dx.doi.org/10.1101/lm.632007.
Coste, C.P., Kleinschmidt, A., 2016. Cingulo-opercular network activity maintains alertness. Neuroimage 128, 264–272. http://dx.doi.org/10.1016/j.neuroimage.2016.01.026.
Critchley, H.D., Tang, J., Glaser, D., Butterworth, B., Dolan, R.J., 2005. Anterior cingulate activity during error and autonomic response. Neuroimage 27 (4), 885–895. http://dx.doi.org/10.1016/j.neuroimage.2005.05.047.
Dawson, M.E., Schell, A.M., Filion, D.L., 2007. The electrodermal system. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (Eds.), Handbook of Psychophysiology2nd ed.. Cambridge University Press, New York, NY, US, 200–223.
De Voogd, L.D., Fernandez, G., Hermans, E.J., 2016a. Disentangling the roles of arousal and amygdala activation in emotional declarative memory. Soc. Cogn. Affect. Neurosci.. http://dx.doi.org/10.1093/scan/nsw055.
De Voogd, L.D., Fernández, G., Hermans, E.J., 2016b. Awake reactivation of emotional memory traces through hippocampal–neocortical interactions. Neuroimage. http://dx.doi.org/10.1016/j.neuroimage.2016.04.026.
Duits, P., Cath, D.C., Lissek, S., Hox, J.J., Hamm, A.O., Engelhard, I.M., Baas, J.M., 2015. Updated meta-analysis of classical fear conditioning in the anxiety disorders. Depress Anxiety 32 (4), 239–253. http://dx.doi.org/10.1002/da.22353.
Ebitz, R.B., Platt, M.L., 2015. Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. Neuron 85 (3), 628–640. http://dx.doi.org/10.1016/j.neuron.2014.12.053.
Eldar, E., Cohen, J.D., Niv, Y., 2013. The effects of neural gain on attention and learning. Nat. Neurosci. 16 (8), 1146–1153. http://dx.doi.org/10.1038/nn.3428.
Etkin, A., Wager, T.D., 2007. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in ptsd, social anxiety disorder, and specific phobia. Am. J. Psychiatry 164 (10), 1476–1488. http://dx.doi.org/10.1176/appi.ajp.2007.07030504.
Etkin, A., Egner, T., Kalisch, R., 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn. Sci. 15 (2), 85–93. http://dx.doi.org/10.1016/j.tics.2010.11.004.
Fullana, M.A., Harrison, B.J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Avila-Parcet, A., Radua, J., 2015. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fmri studies. Mol. Psychiatry.. http://dx.doi.org/10.1038/mp.2015.88.
Gilzenrat, M.S., Nieuwenhuis, S., Jepma, M., Cohen, J.D., 2010. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn. Affect. Behav. Neurosci. 10 (2), 252–269. http://dx.doi.org/10.3758/CABN.10.2.252.
Glover, E.M., Phifer, J.E., Crain, D.F., Norrholm, S.D., Davis, M., Bradley, B., Jovanovic, T., 2011. Tools for translational neuroscience: ptsd is associated with heightened fear responses using acoustic startle but not skin conductance measures. Depress Anxiety 28 (12), 1058–1066. http://dx.doi.org/10.1002/da.20880.
Grady, A.K., Bowen, K.H., Hyde, A.T., Totsch, S.K., Knight, D.C., 2016. Effect of continuous and partial reinforcement on the acquisition and extinction of human conditioned fear. Behav. Neurosci. 130 (1), 36–43. http://dx.doi.org/10.1037/bne0000121.
Grillon, C., Baas, J., 2003. A review of the modulation of the startle reflex by affective states and its application in psychiatry. Clin. Neurophysiol. 114 (9), 1557–1579 , (Retrieved from)⟨http://www.ncbi.nlm.nih.gov/pubmed/12948786⟩.
Grillon, C., Ameli, R., Merikangas, K., Woods, S.W., Davis, M., 1993. Measuring the time course of anticipatory anxiety using the fear-potentiated startle reflex. Psychophysiology 30 (4), 340–346 , (Retrieved from)⟨http://www.ncbi.nlm.nih.gov/pubmed/8327619⟩.
Grillon, C., Lissek, S., Rabin, S., McDowell, D., Dvir, S., Pine, D.S., 2008. Increased anxiety during anticipation of unpredictable but not predictable aversive stimuli as a psychophysiologic marker of panic disorder. Am. J. Psychiatry *165* (7), 898–904. http://dx.doi.org/10.1176/appi.ajp.2007.07101581.
Grillon, C., Pine, D.S., Lissek, S., Rabin, S., Bonne, O., Vythilingam, M., 2009. Increased anxiety during anticipation of unpredictable aversive stimuli in posttraumatic stress disorder but not in generalized anxiety disorder. Biol. Psychiatry 66 (1), 47–53. http://dx.doi.org/10.1016/j.biopsych.2008.12.028.
Hourdaki, E., Giakoumaki, S.G., Grinakis, V., Theou, K., Karataraki, M., Bitsios, P., 2005. Parametric exploration of the fear-inhibited light reflex. Psychophysiology 42 (4), 447–455. http://dx.doi.org/10.1111/j.1469-8986.2005.00301.x.
Joshi, S., Li, Y., Kalwani, Rishi M., Gold, Joshua I., 2016. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. Neuron 89 (1), 221–234. http://dx.doi.org/10.1016/j.neuron.2015.11.028.
Kafkas, A., Montaldi, D., 2011. Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. Q. J. Exp. Psychol. 64 (10), 1971–1989. http://dx.doi.org/10.1080/17470218.2011.588335.
Kaufmann, C., Wehrle, R., Wetter, T.C., Holsboer, F., Auer, D.P., Pollmacher, T., Czisch, M., 2006. Brain activation and hypothalamic functional connectivity during human non-rapid eye movement sleep: an eeg/fmri study. Brain 129 (Pt 3), 655–667. http://dx.doi.org/10.1093/brain/awh686.
Lavin, C., San Martin, R., Rosales Jubal, E., 2014. Pupil dilation signals uncertainty and surprise in a learning gambling task. Front Behav. Neurosci. 7, 218. http://dx.doi.org/10.3389/fnbeh.2013.00218.
Liddell, B.J., Brown, K.J., Kemp, A.H., Barton, M.J., Das, P., Peduto, A., Williams, L.M., 2005. A direct brainstem-amygdala-cortical 'alarm' system for subliminal signals of fear. Neuroimage 24 (1), 235–243. http://dx.doi.org/10.1016/j.neuroimage.2004.08.016.
Lindner, K., Neubert, J., Pfannmoller, J., Lotze, M., Hamm, A.O., Wendt, J., 2015. Fear-potentiated startle processing in humans: parallel fmri and orbicularis emg assessment during cue conditioning and extinction. Int. J. Psychophysiol. 98 (3 Pt 2), 535–545. http://dx.doi.org/10.1016/j.ijpsycho.2015.02.025.
Linnman, C., Rougemont-Bucking, A., Beucke, J.C., Zeffiro, T.A., Milad, M.R., 2011. Unconditioned responses and functional fear networks in human classical conditioning. Behav. Brain Res. 221 (1), 237–245. http://dx.doi.org/10.1016/j.bbr.2011.02.045.
Lipp, O.V., Sheridan, J., Siddle, D.A., 1994. Human blink startle during aversive and nonaversive pavlovian conditioning. J. Exp. Psychol. Anim. Behav. Process 20 (4), 380–389. http://dx.doi.org/10.1037/0097-7403.20.4.380.
Lissek, S., Baas, J.M., Pine, D.S., Orme, K., Dvir, S., Nugent, M., Grillon, C., 2005a. Airpuff startle probes: an efficacious and less aversive alternative to white-noise. Biol. Psychol. 68 (3), 283–297. http://dx.doi.org/10.1016/j.biopsycho.2004.07.007.
Lissek, S., Powers, A.S., McClure, E.B., Phelps, E.A., Woldehawariat, G., Grillon, C., Pine, D.S., 2005b. Classical fear conditioning in the anxiety disorders: a meta-analysis. Behav. Res Ther. 43 (11), 1391–1424. http://dx.doi.org/10.1016/j.brat.2004.10.007.
Lissek, S., Biggs, A.L., Rabin, S.J., Cornwell, B.R., Alvarez, R.P., Pine, D.S., Grillon, C., 2008a. Generalization of conditioned fear-potentiated startle in humans: experimental validation and clinical relevance. Behav. Res Ther. 46 (5), 678–687. http://dx.doi.org/10.1016/j.brat.2008.02.005.
Lissek, S., Levenson, J., Biggs, A.L., Johnson, L.L., Ameli, R., Pine, D.S., Grillon, C., 2008b. Elevated fear conditioning to socially relevant unconditioned stimuli in social anxiety disorder. Am. J. Psychiatry 165 (1), 124–132. http://dx.doi.org/10.1176/appi.ajp.2007.06091513.
Lissek, S., Rabin, S., Heller, R.E., Lukenbaugh, D., Geraci, M., Pine, D.S., Grillon, C., 2010. Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. Am. J. Psychiatry 167 (1), 47–55. http://dx.doi.org/10.1176/appi.ajp.2009.09030410.
Lowenstein, O., Feinberg, R., Loewenfeld, I.E., 1963. Pupillary movements during acute and chronic fatigue – a new test for the objective evaluation of tiredness. Investig. Ophthalmol. 2 (2), 138–157.
Maier, S., Szalkowski, A., Kamphausen, S., Perlov, E., Feige, B., Blechert, J., Tuscher, O., 2012. Clarifying the role of the rostral dmpfc/dacc in fear/anxiety: learning, appraisal or expression? PLoS One 7 (11), e50120. http://dx.doi.org/10.1371/journal.pone.0050120.
Manohar, S.G., Husain, M., 2016. Human ventromedial prefrontal lesions alter incentivisation by reward. Cortex 76, 104–120. http://dx.doi.org/10.1016/j.cortex.2016.01.005.
Mechias, M.L., Etkin, A., Kalisch, R., 2010. A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. Neuroimage 49 (2), 1760–1768. http://dx.doi.org/10.1016/j.neuroimage.2009.09.040.
Milad, M.R., Quirk, G.J., 2012. Fear extinction as a model for translational neuroscience: ten years of progress. Annu. Rev. Psychol. 63, 129–151. http://dx.doi.org/10.1146/annurev.psych.121208.131631.
Milad, M.R., Quirk, G.J., Pitman, R.K., Orr, S.P., Fischl, B., Rauch, S.L., 2007. A role for the human dorsal anterior cingulate cortex in fear expression. Biol. Psychiatry 62 (10), 1191–1194. http://dx.doi.org/10.1016/j.biopsych.2007.04.032.
Morriss, J., Christakou, A., van Reekum, C.M., 2015. Intolerance of uncertainty predicts fear extinction in amygdala-ventromedial prefrontal cortical circuitry. Biol. Mood Anxiety Disord. 5, 4. http://dx.doi.org/10.1186/s13587-015-0019-8.
Murphy, P.R., O'Connell, R.G., O'Sullivan, M., Robertson, I.H., Balsters, J.H., 2014. Pupil diameter covaries with bold activity in human locus coeruleus. Hum. Brain Mapp. 35 (8), 4140–4154. http://dx.doi.org/10.1002/hbm.22466.
Nassar, M.R., Rumsey, K.M., Wilson, R.C., Parikh, K., Heasly, B., Gold, J.I., 2012. Rational regulation of learning dynamics by pupil-linked arousal systems. Nat. Neurosci. 15 (7), 1040–1046. http://dx.doi.org/10.1038/nn.3130.
Nieuwenhuis, S., De Geus, E.J., Aston-Jones, G., 2011. The anatomical and functional relationship between the p3 and autonomic components of the orienting response. Psychophysiology 48 (2), 162–175. http://dx.doi.org/10.1111/j.1469-8986.2010.01057.x.
Öhman, A., 1971. Differentiation of conditioned and orienting response components in electrodermal conditioning. Psychophysiology 8 (1), 7–22 , (Retrieved from)⟨http://www.ncbi.nlm.nih.gov/pubmed/5578384⟩.
Öhman, A., 2005. The role of the amygdala in human fear: automatic detection of threat. Psychoneuroendocrinology 30 (10), 953–958. http://dx.doi.org/10.1016/j.psyneuen.2005.03.019.
Partala, T., Surakka, V., 2003. Pupil size variation as an indication of affective processing. Int. J. Hum. – Comput. Stud. 59 (1-2), 185–198. http://dx.doi.org/10.1016/s1071-5819(03)00017-x.
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. http://dx.doi.org/10.1111/j.1469-8986.2009.00852.x.
Pole, N., 2007. The psychophysiology of posttraumatic stress disorder: a meta-analysis. Psychol. Bull. 133 (5), 725–746. http://dx.doi.org/10.1037/0033-2909.133.5.725.
Pollak, D.D., Rogan, M.T., Egner, T., Perez, D.L., Yanagihara, T.K., Hirsch, J., 2010. A

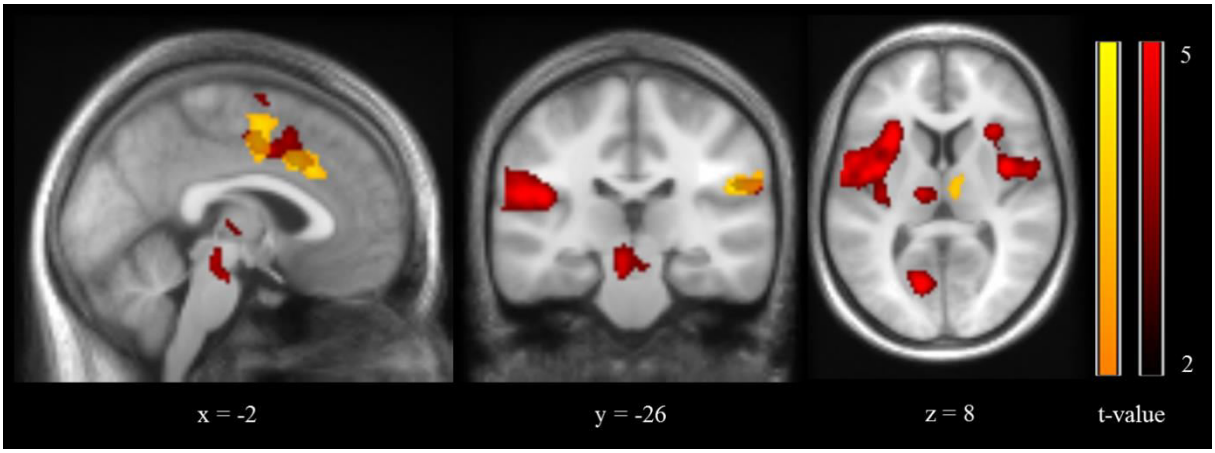
translational bridge between mouse and human models of learned safety. Ann. Med. 42 (2), 115–122. http://dx.doi.org/10.3109/07853890903583666.
Preuschoff, K., Hart, B.M., Einhauser, W., 2011. Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. Front. Neurosci. 5, 115. http://dx.doi.org/10.3389/fnins.2011.00115.
Reinhard, G., Lachnit, H., 2002. Differential conditioning of anticipatory pupillary dilation responses in humans. Biol. Psychol. 60 (1), 51–68. http://dx.doi.org/10.1016/S0301-0511(02)00011-X.
Reinhard, G., Lachnit, H., Konig, S., 2006. Tracking stimulus processing in pavlovian pupillary conditioning. Psychophysiology 43 (1), 73–83. http://dx.doi.org/10.1111/j.1469-8986.2006.00374.x.
Sadaghiani, S., D'Esposito, M., 2015. Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. Cereb. Cortex 25 (9), 2763–2773. http://dx.doi.org/10.1093/cercor/bhu072.
Samuels, E.R., Szabadi, E., 2008. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part i: principles of functional organisation. Curr. Neuropharmacol. 6 (3), 235–253. http://dx.doi.org/10.2174/157015908785777229.
Schneider, M., Hathway, P., Leuchs, L., Samann, P.G., Czisch, M., Spoormaker, V.I., 2016. Spontaneous pupil dilations during the resting state are associated with activation of the salience network. Neuroimage. http://dx.doi.org/10.1016/j.neuroimage.2016.06.011.
Sears, R.M., Fink, A.E., Wigestrand, M.B., Farb, C.R., de Lecea, L., Ledoux, J.E., 2013. Orexin/hypocretin system modulates amygdala-dependent threat learning through the locus coeruleus. Proc. Natl. Acad. Sci. USA 110 (50), 20260–20265. http://dx.doi.org/10.1073/pnas.1320325110.
Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Greicius, M.D., 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27 (9), 2349–2356. http://dx.doi.org/10.1523/Jneurosci.5587-06.2007.
Sehlmeyer, C., Schoning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., Konrad, C., 2009. Human fear conditioning and extinction in neuroimaging: a systematic review. PLoS One 4 (6), e5865. http://dx.doi.org/10.1371/journal.pone.0005865.
Seifert, F., Schuberth, N., De Col, R., Peltz, E., Nickel, F.T., Maihofner, C., 2013. Brain activity during sympathetic response in anticipation and experience of pain. Hum. Brain Mapp. 34 (8), 1768–1782. http://dx.doi.org/10.1002/hbm.22035.
Siegle, G.J., Steinhauer, S.R., Stenger, V.A., Konecky, R., Carter, C.S., 2003. Use of concurrent pupil dilation assessment to inform interpretation and analysis of fmri data. Neuroimage 20 (1), 114–124. http://dx.doi.org/10.1016/S1053-8119(03)00298-2.
Sirois, S., Brisson, J., 2014. Pupillometry. Wiley Interdiscip. Rev. Cogn. Sci. 5 (6), 679–692. http://dx.doi.org/10.1002/wcs.1323.
Spoormaker, V.I., Blechert, J., Goya-Maldonado, R., Samann, P.G., Wilhelm, F.H., Czisch, M., 2012. Additional support for the existence of skin conductance responses at unconditioned stimulus omission. Neuroimage 63 (3), 1404–1407. http://dx.doi.org/10.1016/j.neuroimage.2012.08.050.
Spoormaker, V.I., Andrade, K.C., Schroter, M.S., Sturm, A., Goya-Maldonado, R., Samann, P.G., Czisch, M., 2011. The neural correlates of negative prediction error signaling in human fear conditioning. Neuroimage 54 (3), 2250–2256. http://dx.doi.org/10.1016/j.neuroimage.2010.09.042.
Sterpenich, V., D'Argembeau, A., Desseilles, M., Balteau, E., Albouy, G., Vandewalle, G., Maquet, P., 2006. The locus ceruleus is involved in the successful retrieval of emotional memories in humans. J. Neurosci. 26 (28), 7416–7423. http://dx.doi.org/10.1523/jneurosci.1001-06.2006.
Vaidyanathan, U., Patrick, C.J., Cuthbert, B.N., 2009. Linking dimensional models of internalizing psychopathology to neurobiological systems: affect-modulated startle as an indicator of fear and distress disorders and affiliated traits. Psychol. Bull. 135 (6), 909–942. http://dx.doi.org/10.1037/a0017222.
Van Well, S., Visser, R.M., Scholte, H.S., Kindt, M., 2012. Neural substrates of individual differences in human fear learning: evidence from concurrent fmri, fear-potentiated startle, and us-expectancy data. Cogn. Affect Behav. Neurosci. 12 (3), 499–512. http://dx.doi.org/10.3758/s13415-012-0089-7.
Visser, R.M., Scholte, H.S., Beemsterboer, T., Kindt, M., 2013. Neural pattern similarity predicts long-term fear memory. Nat. Neurosci. 16 (4), 388–390. http://dx.doi.org/10.1038/nn.3345.
Visser, R.M., Kunze, A.E., Westhoff, B., Scholte, H.S., Kindt, M., 2015. Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. Psychoneuroendocrinology 55, 8–20. http://dx.doi.org/10.1016/j.psyneuen.2015.01.021.
Visser, R.M., de Haan, M.I., Beemsterboer, T., Haver, P., Kindt, M., Scholte, H.S., 2016. Quantifying learning-dependent changes in the brain: single-trial multivoxel pattern analysis requires slow event-related fmri. Psychophysiology 53 (8), 1117–1127. http://dx.doi.org/10.1111/psyp.12665.
Wang, C.A., Munoz, D.P., 2015. A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. Curr. Opin. Neurobiol. 33, 134–140. http://dx.doi.org/10.1016/j.conb.2015.03.018.
Wiemer, J., Muhlberger, A., Pauli, P., 2014. Illusory correlations between neutral and aversive stimuli can be induced by outcome aversiveness. Cogn. Emot. 28 (2), 193–207. http://dx.doi.org/10.1080/02699931.2013.809699.
Wilhelm, H., Ludtke, H., Wilhelm, B., 1998. Pupillographic sleepiness testing in hypersomniacs and normals. Graefes Arch. Clin. Exp. Ophthalmol. 236 (10), 725–729. http://dx.doi.org/10.1007/s004170050149.
Zekveld, A.A., Heslenfeld, D.J., Johnsrude, I.S., Versfeld, N.J., Kramer, S.E., 2014. The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. Neuroimage 101, 76–86. http://dx.doi.org/10.1016/j.neuroimage.2014.06.069.

6. Supplementary Material

6.1 Supplementary Figures

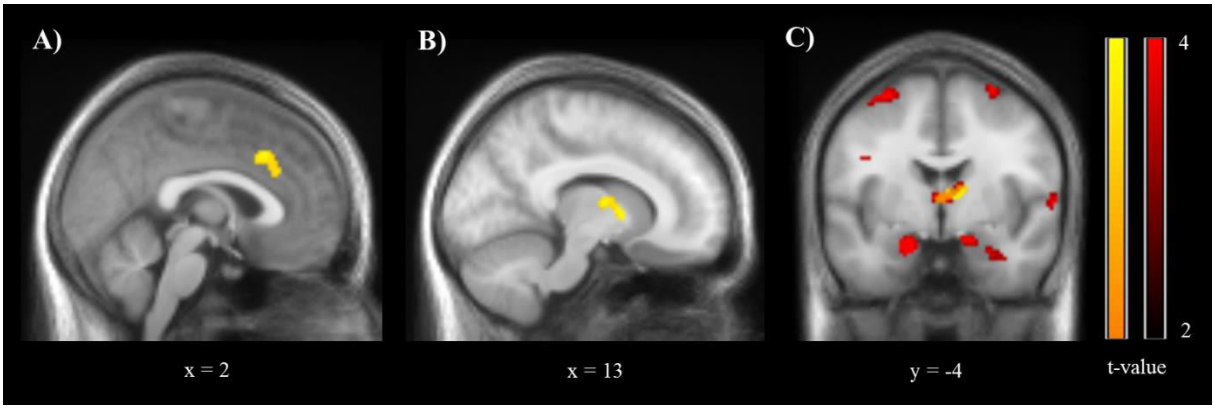


Supplementary Figure S1. Baseline pupil diameter (average over 0.5 s pre-stimulus) for all 46 trials throughout the fear learning task (N = 26).



Supplementary Figure S2. Activity associated with the stimulus contrast CS- < [CS60+ and CS100+] throughout the whole fear learning task in red (cluster extension threshold set to k = 79), derived from a full-factorial ANOVA with the levels CS-, CS60+ and CS100+; (contrast [-2 +1 +1], N = 34). Overlay in yellow: activity positively correlated with the magnitude of per-trial pupil responses within stimulus types throughout the whole fear learning task (cluster extension threshold set to k = 108), derived from a full-factorial ANOVA of the parametric modulations to CS-, CS60+ and CS100+ in separate regressors; (contrast [+1 +1 +1], N =

26). Maps were sampled at $p < .001$ (uncorrected) with a whole-brain corrected cluster threshold of $p_{FWE} < .05$. Coordinates [mm] refer to MNI space.



Supplementary Figure S3. A) Activity positively correlated with pupil response magnitude during fear conditioning (cluster extension threshold set to k = 18, N = 28). **B)** Activity positively correlated with baseline-weighted pupil responses during fear conditioning (cluster extension threshold set to k = 68, N = 28). Both A) and B) were derived from a full-factorial ANOVA containing parametric modulations to CS-, CS60+ and CS100+ in separate regressors; (contrasts [+1 +1 +1]). **C)** Negative effect of a linear time modulation to the three stimuli throughout fear conditioning trials in red (cluster extension threshold set to k = 90). The post-hoc t-contrast was derived from a full-factorial ANOVA with the parametric modulations to CS-, CS60+ and CS100 in separate regressors (contrast [-1 -1 -1], N = 34). Overlay with the thalamic cluster from B) in yellow. Maps were sampled at $p < .001$ (uncorrected) with a whole-brain corrected cluster threshold of $p_{FWE} < .05$. Coordinates [mm] refer to MNI space.

6.2 Supplementary Tables

Supplementary Table S1. Activity associated with the stimulus contrast CS- < [CS60+ and CS100+] during fear conditioning (N = 34). The post-hoc t-contrast [-2 +1 +1] was derived from a full-factorial ANOVA with the levels CS-, CS60+ and CS100+. Maps were sampled at $p < .001$ (uncorrected) with a whole-brain corrected cluster threshold of $p_{FWE} < .05$.

Region name	Cluster		Voxel			
	P_{corr}	k	T_{peak}	MNI_{peak} x	y	z
Insula (L) Superior Temporal gyrus (L) Rolandic operculum (L) Supramarginal gyrus (L) Postcentral gyrus (L)	< .001*	5388	7.50	-44	-26	16
Calcarine sulcus (L+R) Lingual gyrus (L+R) Cuneus (L+R) Vermis (lobule 6)	< .001*	2189	6.71	-12	-70	8
Insula (R) Rolandic operculum (R) Superior temporal gyrus (R) Supramarginal gyrus (R)	< .001*	3637	6.36	38	0	10
Supplementary motor area (L+R) Precentral gyrus (L) Middle cingulate cortex (L+R) Postcentral gyrus (L) Paracentral lobule (L)	< .001*	5456	6.09	-8	16	38
Thalamus (L+R) Amygdala (R, 3.4% of cluster)	< .001*	1293	5.84	-4	-24	-20
Precentral gyrus (R) Middle frontal gyrus (R)	.013	190	4.44	50	-4	50

Supplementary Table S2. Activity associated with the stimulus contrast CS- > [CS60+ and CS100+] during fear conditioning (N = 34). The post-hoc t-contrast [+2 -1 -1] was derived from a full-factorial ANOVA with the levels CS-, CS60+ and CS100+. Maps were sampled at $p < .001$ (uncorrected) with a whole-brain corrected cluster threshold of $p_{FWE} < .05$.

	Cluster		Voxel			
Region name	**P_{corr}**	**k**	**T_{peak}**	**MNI_{peak}**		
				x	**y**	**z**
Middle temporal gyrus (L)	.046*	141	5.20	-60	-14	-16
Superior medial frontal gyrus (L+R) Superior frontal gyrus (L+R)	.002	281	4.49	8	60	32
Medial orbitofrontal cortex (L+R) Gyrus rectus (L) Superior medial frontal gyrus (L)	.027	161	4.15	-2	48	-18

Note: Supplementary Tables S1 and S2 refer to Figure 3. Maps were sampled at $p < .001$ (uncorrected) with a whole-brain corrected cluster threshold of $p_{FWE} < .05$. Only brain regions contributing over 5% to respective clusters are listed. R = right; L = left; P_{corr} stands for whole brain corrected cluster p-values, k for the cluster size, t_{peak} for the t-value of the peak-voxel, [x y z] coordinates are in MNI-space.

Supplementary Table S3. Activity negatively associated with the (increasing) linear time modulation to all three stimuli during fear conditioning trials. The post-hoc t-contrast was derived from a full-factorial ANOVA with the linear time modulations to CS-, CS60+ and CS100+ in separate regressors (contrast [-1 -1 -1], N = 34). Maps were sampled at $p < .001$ (uncorrected) with a whole-brain corrected cluster threshold of $p_{FWE} < .05$.

Region name	Cluster		Voxel			
	P_{corr}	k	T_{peak}	MNI_{peak} x	y	z
Inferior and middle Temporal gyrus (R)	.013*	168	6.31	48	-46	-8
Superior and middle frontal gyrus (R)	< .001*	339	5.71	20	-16	60
Precentral gyrus (L) Postcentral gyrus (L) Middle frontal gyrus (L)	< .001*	956	5.64	-46	-22	58
Parahippocampal gyrus (R) Hippocampus (R) Lingual gyrus (R)	.002*	249	5.45	24	-16	-20
Calcarine sulcus (L+R) Lingual gyrus (L) Lobule 4 &5 of vermis (L) Precuneus (L)	< .001*	621	5.43	-2	-52	4
Hippocampus (L) Fusiform gyrus (L) Parahippocampal gyrus (L) Lobule 4 & 5 of cerebelum (L) Amygdala (4.71% of cluster, L)	< .001*	1104	5.24	-20	-10	-18
Superior temporal gyrus (L) Postcentral gyrus (L)	< .001*	838	5.20	-56	-30	10
Superior and inferior parietal lobule (R) Postcentral gyrus (R) Supramarginal gyrus (R)	< .001	445	5.17	38	-40	50
Lobule 4, 5 & 6 of cerebelum (R) Lingual gyrus (R) Fusiform gyrus (R)	< .001	594	4.96	20	-40	-32
Amygdala (R) Parahippocampal gyrus (R) Hippocampus (R) Fusiform gyrus (R)	.005	206	4.95	20	0	-18
Thalamus (L+R)	.011	173	4.90	0	-8	6
Precuneus (L)	< .001	388	4.78	-14	-42	64

Paracentral lobule (L+R) Postcentral gyrus (R)						
Superior temporal gyrus (R) Rolandic operculum (R) Transverse temporal gyrus (R)	< .001	384	4.77	64	-10	6
Inferior and middle temporal gyrus (L) Crus 1 of celebelum (L)	< .001	324	4.72	-44	-70	-26
Postcentral gyrus (L) Superior and inferior parietal lobule (L)	.026	144	4.59	-36	-36	66
Middle occipital gyrus (R)	.031	138	4.45	38	-72	26
Middle and superior occipital gyrus (L)	.008	187	4.26	-24	-80	40

Note: Supplementary Table S3 refers to Supplementary Figure S3C. Maps were sampled at $p < .001$ (uncorrected) with a whole-brain corrected cluster threshold of $p_{FWE} < .05$. Only brain regions contributing over 5% to respective clusters are listed. R = right; L = left; P_{corr} stands for whole brain corrected cluster p-values, k for the cluster size, t_{peak} for the t-value of the peak-voxel, [x y z] coordinates are in MNI-space.

4. Discussion

4.1 Cognitive-affective processes influencing physiological measures during fear learning

Psychophysiological measures allow only indirect inferences about ongoing cognitive processes and related brain activity patterns. Neither the pupillary muscles nor the orbicularis oculi (startle EMG) or eccrine sweat glands (SCR) receive direct projections from higher-order brain areas. Therefore, cognitive processes need to be mediated via the autonomous nervous system to result in physiological changes. Given this indirect and rather unspecific association of the physiological periphery with cortical activity, the experimental contexts, as well as the properties of specific physiological readouts are decisive for the interpretation of physiological responses.

4.1.1 Slow pupil dilations

Presentations of conditioned stimuli elicited changes in pupil diameter. We observed large, reflexive changes at CS onset and offset. These were common to all stimulus presentations and were due to the differing screen brightness between stimuli and inter-trial intervals. By contrast, slow pupil dilations throughout the stimulus interval differed between stimuli: in both studies, they displayed larger dilations in response to CS+ than CS- at the end of fear acquisition. These differences should be attributable to the influence of ongoing cognitive-affective processes. As stated in the introduction, the pupil has previously been shown to respond to emotional material (Bradley et al., 2008; Partala & Surakka, 2003; Snowden et al., 2016). In our studies, we used emotionally neutral cues as conditioned stimuli. However, CS+ and CS- may acquire a negative or positive emotional valence throughout the fear learning process due to their association with the aversive US or safety, respectively. Besides the acquired emotional valence of the conditioned stimuli, declarative learning of the CS-US contingencies and US expectancy might involve other cognitive processes. For example, it has been shown that the pupil is sensitive to outcome prediction, error monitoring and surprise, cognitive effort or uncertainty (Alnaes et al., 2014; Jepma & Nieuwenhuis, 2011; Lavin et al., 2014; Preuschoff et al., 2011; Wendt et al., 2016), which may additionally contribute to pupil dilation during fear learning.

US expectancy

Our results indicate that slow pupil dilations in large parts relate to threat appraisal and explicit US expectancy for several reasons. First, pupil dilations roughly followed US expectancy ratings in both studies. US expectancy ratings reflected on average the actual reinforcement rates, indicating that subjects have explicitly learned the CS-US contingencies and likely expected the US at CS+ offset with rather accurate

probabilities. Pupil dilations mirrored these ratings: they initially increased in response to CS+ in both studies, and pupil dilations were initially stronger in response to a fully reinforced CS+ than to a partially reinforced CS+ in the first half of fear conditioning (in the MR study). Second, pupil dilations discriminated maximally between CS+ and CS- in the time window immediately preceding the US (demonstrated in the MR study as well as by Reinhard & Lachnit, 2002). The difference in pupil dilations to CS+ and CS- therefore increases throughout the stimulus interval and is maximal in direct outcome anticipation. This suggests that explicit US expectancy drives a large part of the stimulus difference in pupil dilations. A third indicator for this is the detected close association of pupil dilations with dACC activity. This region has previously been hypothesized to reflect conscious threat appraisal during fear conditioning (Etkin, Egner, & Kalisch, 2011; Maier et al., 2012; Mechias et al., 2010). The association of pupil dilations and dACC will be discussed in further detail below.

Uncertainty

Our results suggest that pupil dilations do not only reflect US expectancy, but other processes involved in declarative fear learning. These involve uncertainty or prediction error signaling, which have previously been found to affect pupil dilations (Lavin et al., 2014; Morriss et al., 2015; Nassar et al., 2012; Preuschoff et al., 2011). These processes may influence pupil dilations in particular if intermittent reinforcement schedules are employed, i.e. when the occurrence of a US at CS offset is uncertain and difficult to predict. Another circumstance which may trigger this process is a change of reinforcement rules, for example at the beginning of extinction learning (i.e. when an expected US does not occur). Both our fear learning tasks comprised intermittent reinforcement as well as extinction learning, therefore updating of US expectancy and uncertainty about the CS outcome may have contributed to pupil dilations.

Koenig et al. (2017) provided data suggesting that anticipatory pupil dilations specifically reflected prediction-error driven outcome expectancy in their conditioning tasks. They found a consistent pattern during both aversive and appetitive conditioning: during early learning stages fully reinforced cues elicited greater pupil dilation than intermittently reinforced cues. This pattern reversed in later learning stages, in which cues with an uncertain outcome elicited stronger pupil dilations. Our observations were very similar for the MR study, where we employed two CS+ with different reinforcement rates (60% and 100% reinforcement). In early stages of fear acquisition, the fully reinforced CS+ elicited stronger pupil dilations than the partially reinforced CS+ (see Figure 2 A). In this task stage, pupil responses matched US expectancy ratings and thereby seemed to roughly track US expectancy. In later task stages, pupil dilations diverged from US expectancy ratings. They declined in response to the fully reinforced CS+ and later converged with pupil responses to the partially reinforced CS+, which roughly remained at the same level throughout fear acquisition. Similarly, slow pupil dilations stayed elevated in response to CS+ during extinction (in the MR experiment) as well as during recall (in the laboratory experiment), whereas US expectancy ratings

decreased. In later task stages, pupil dilations might hence primarily reflect prolonged uncertainty about the outcome of the previously reinforced CS, which is why they diverge from US expectancy ratings. Our results therefore support the notion that expectancy updating and uncertainty about the CS outcome contribute to slow pupil dilations in response to conditioned stimuli (Koenig et al., 2017).

We found no evidence for a partial reinforcement extinction effect (PREE) in our MR study, i.e. pupil responses during extinction were not larger to partially reinforced CS+ than to fully reinforced CS+. In line with models on reinforcement learning (e.g. Rescorla & Wagner, 1972), it was to be anticipated that extinction would occur more quickly for a previously fully reinforced stimulus (with larger prediction errors and a higher learning rate). Given the demonstrated sensitivity of pupil dilations to US expectancy and uncertainty during fear learning, the lack of a PREE might have been due to the rather small sample size (N~14 per group) and the between-subjects design, or due to a generally small PREE effect. Overall, our results indicate that reinforcement rates as well as the learning stage are decisive for the interpretation of pupil dilations in terms of threat appraisal and explicit US expectancy (early stages) or in terms of CS outcome uncertainty (pointing towards cumulative, error-driven responses for partial reinforcement in later learning stages, Koenig et al., 2017; Pearce & Hall, 1980).

4.1.2 Reflexive pupil responses

Reflexive pupillary readouts, such as auditory pupil responses or the pupillary light reflex, may differ from slow pupil dilations. The major difference of triggered responses is that they are provoked by a probe during the presence of a CS and not elicited by the CS itself. Furthermore, the temporal scale of such reflexive responses is much shorter (1 s in our case as opposed to 4 s), which likely prevents a strong influence of higher cognitive processes on the response magnitude. The auditory pupil reflex to startle probes yielded significant stimulus discrimination during fear acquisition, with larger responses to startle probes when a CS+ was present (as opposed to during CS- or inter-trial intervals).

To date, pupil responses to startle probes have not been published as a measure of the conditioned response, but pupillary reflexes to sound material have been investigated before (e.g. Marois et al., 2017; Wetzel, Buttelmann, Schieler, & Widmann, 2016; Widmann, Schroger, & Wetzel, 2018). It has previously been suggested that the initial pupil reflex to sounds may display properties of an orienting response, which is modulated by the novelty and saliency of the stimulus material (MacDonald & Barry, 2017; Sokolov, 1963, 1990). Some studies argue for the notion that pupil dilations in fact reflect orienting by showing increased reflexive responses to deviant sounds and visual contrast, i.e. to stimulus novelty and saliency (Marois et al., 2017; Wang, Boehnke, Itti, & Munoz, 2014). Steiner and Barry (2011) argue against this account, showing that initial pupil responses were only modulated by stimulus novelty, but not by stimulus 'significance' for the task. This conflict might be resolved by differentiating between 'saliency' in terms of inherent stimulus properties (e.g. visual contrast, loudness or pitch) or in terms of task-relevance and emotional meaning, which

requires an interpretation of the stimulus material. In this context, it is also crucial to take into account which temporal window of the pupillary response is considered. The very initial pupil dilation response (within approximately 1 s) has been demonstrated to solely relate to stimulus novelty (Geva, Zivan, Warsha, & Olchik, 2013; Steiner & Barry, 2011; Widmann et al., 2018) and to inherent stimulus properties (Marois et al., 2017; Wang et al., 2014), but not to stimulus saliency in terms of its emotional relevance (Steiner & Barry, 2011; Widmann et al., 2018). Widmann et al. (2018) proposed that the pupil response can be split into two temporal components which reflect parasympathetic inhibition and sympathetic activation, respectively. Both processes can result in pupil dilation, however the prior disappears in darkness: in this state the parasympathetic constrictor muscles are maximally relaxed and parasympathetic inhibition does not affect the pupil diameter. This way, Widmann et al. (2018) demonstrated that the initial pupil response is driven by parasympathetic inhibition and is only sensitive to the novelty but not the emotional content (i.e. saliency) of the stimulus material. A later component (initiated after approximately 1 s) was enhanced by emotionally arousing sounds, i.e. sensitive to the saliency of the stimulus material and was suggested to reflect sympathetic activation. An emotional evaluation of stimulus saliency therefore seems to be restricted to later pupil dilations and appears precluded from initial, reflexive pupil responses. In line with this, the initial light reflex in response to CS onset was not modulated by fear as observed in our MR study, whereas slow pupil dilations later throughout the stimulus interval were.

Another question is whether the detected auditory pupil reflex to the startle probes can be explained in terms of an orienting response in our study, but we argue against that. First, the auditory pupil reflex did not habituate with declining stimulus (i.e. startle probe) novelty. Second, despite displaying differences between CS+ and CS-, this difference cannot be related to startle probe saliency: the response-eliciting stimulus was always the same startle probe for both CS+ and CS-. Here, we propose that elicited pupillary reflexes during fear conditioning are not modulated by differential orienting to the probe, but by the underlying state of physiological arousal at the time of probing. This interpretation can also be extrapolated to the fear-inhibited light reflex, which is elicited by neutral light probes during the stimulus interval (Bitsios, Szabadi, & Bradshaw, 1996; Bitsios et al., 2004). We triggered auditory pupil responses during states that potentially differ in their arousal level, i.e. during late stages of the CS+ and CS- stimulus intervals and during inter-trial intervals. Physiological arousal itself can in turn be determined by different factors: first, US expectancy and threat appraisal are likely to contribute to physiological arousal, as auditory pupil responses were enhanced during CS+ presentations and increased throughout fear learning (CS+ > CS- difference, similar to slow pupil dilations). Second, attentional processes might contribute to physiological arousal and response magnitude, resulting in enhanced responses to startle probes during stimulus intervals as opposed to inter-trial intervals (CS- > inter-trial interval).

4.1.3 Tonic pupil diameter: overall arousal / wakefulness:

We found slow pupil dilations as well as auditory pupil responses to be unaffected by physiological habituation, which stands in contrast to SCR and startle responses. Contrarily, the baseline pupil diameter declined in all datasets (in the MR and laboratory environment) and throughout all task phases (during fear acquisition, extinction as well as recall). It has previously been suggested that tonic changes in pupil diameter inform about changes in wakefulness and overall arousal (Aston-Jones & Cohen, 2005; Lowenstein, Feinberg, & Loewenfeld, 1963; Wilhelm, Ludtke, & Wilhelm, 1998).

In each dataset, we reliably found a moderate, negative association of the pupil baseline diameter with the superimposed phasic pupil dilations. This was apparent for slow pupil dilations and, to a smaller extent, for auditory pupil responses. This negative relationship has been reported before in macaque monkeys (Joshi et al., 2016) and in humans (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010). The reason for this phenomenon is not yet known: it could be explained by simple ceiling or floor effects, i.e. there may be less margin for phasic dilations if the pupil is already strongly dilated. Similarly, the pupil is limited in its constriction. This interdependence suggests that the lack of habituation in phasic pupil responses has to be considered with caution: pupil responses might actually habituate over time. This might however not become apparent due to the constant decrease in tonic baseline diameter, which may in turn be causal for simultaneously increasing phasic response magnitude.

Our findings demonstrate that the tonic pupil diameter may convey complementary information to phasic responses throughout fear learning and that it cannot necessarily be considered as independent of phasic responses. To date, some studies have adjusted phasic responses for tonic changes in pupil diameter by division or multiplication (De Voogd, Fernández, & Hermans, 2016; Marois et al., 2017), while most fear conditioning studies have subtracted the preceding baseline diameter from phasic responses (e.g. Reinhard & Lachnit, 2002; Reinhard et al., 2006; Visser et al., 2016; Visser et al., 2015; Visser et al., 2013). Here we suggest that the tonic pupil diameter should always be taken into account, or at least be reported upon interpreting phasic pupil dilations.

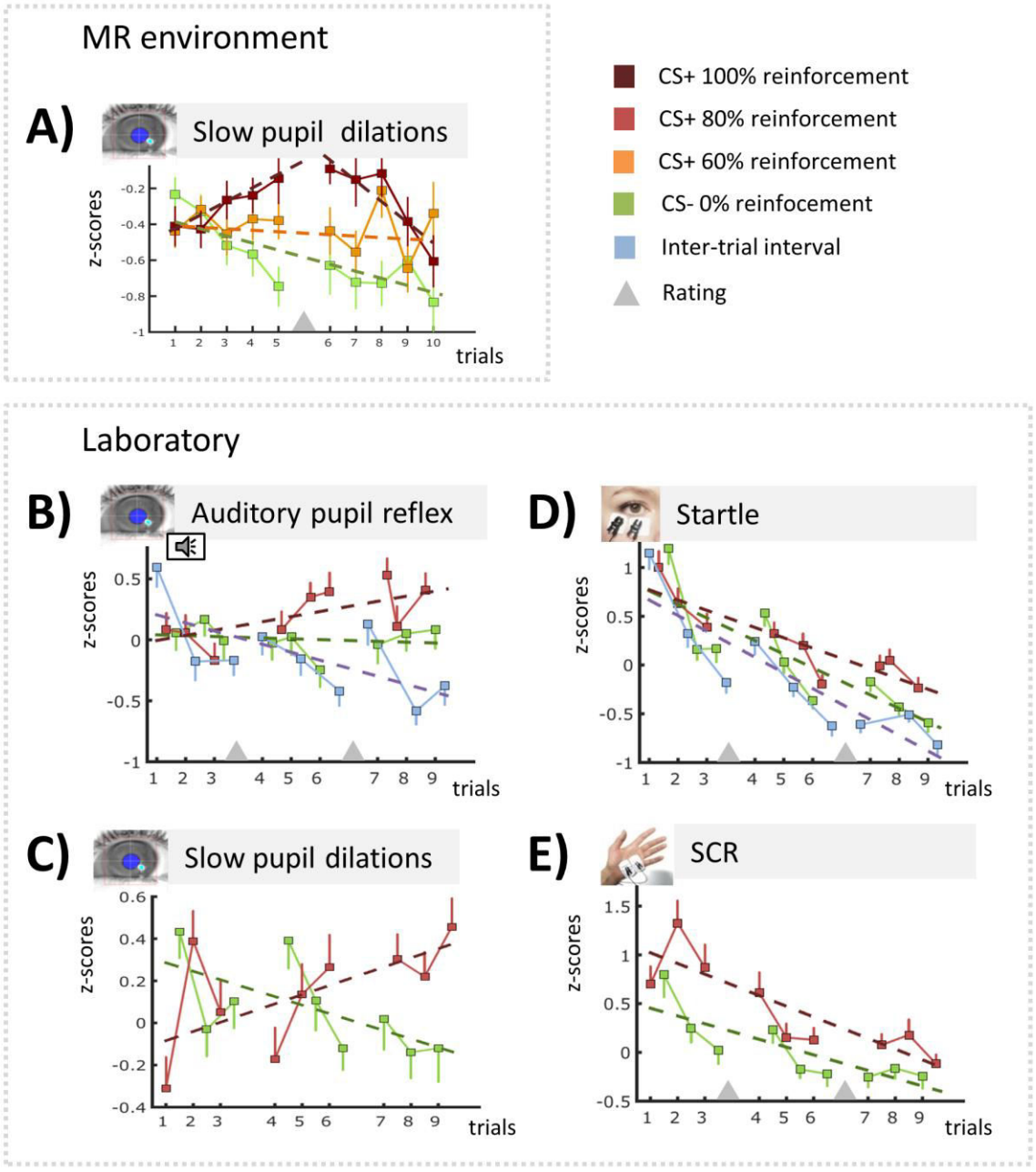


Figure 2. Physiological responses throughout fear acquisition from the MR study (A) and the laboratory study (B-E). In the MR environment, two CS+ were reinforced in 60% and 100% of all trials, respectively. In the laboratory, CS+ was reinforced in 80% of all trials. Responses to different stimuli (CS or startle probes) are marked in color: responses to CS- (or startle probes applied during CS- in B and D) in green; responses to CS+ (or startle probes applied during CS+ in B and D) in red (80% and 100% reinforcement) and orange (60% reinforcement); responses to inter-trials interval startle probes in blue (in B and D). Ratings of US probability are marked by grey triangles. Vertical bars reflect the standard error of the mean (SEM).

4.1.4 Relating pupil responses to SCR and startle responses

Habituation

In our laboratory study, we found that the major difference between pupillometric and other measures consisted in their response patterns over time: both SCR and startle responses were strongly affected by habituation and declined in response to both CS- and CS+ (see Figure 2 D & E). Groves and Thompson (1970) define habituation as a "decremental process" in the stimulus-response pathway, which is inversely related to stimulus intensity and directly related to stimulus frequency (i.e. habituation increases with the amount of stimulus repetitions). SCR and startle responses therefore likely decline due to habituation, caused by repetitive presentations to the conditioned stimuli or startle probes (at a constant stimulus intensity). Physiological habituation is hence a major factor contributing to the magnitude of SCR and startle response, but not to pupil dilations.

Valence

Whereas habituation affected SCR and startle responses throughout all task phases, other cognitive affective processes are likely responsible for the detected CS+ > CS- stimulus difference during fear acquisition. One such factor might be an acquired negative valence of the CS+. As stated above, the CS+ likely acquires negative emotional valence with progressing fear acquisition, due to its repetitive association with the aversive shock. We cannot draw direct conclusions on the valence-specificity of either physiological measure from our studies, because we did not explicitly contrast stimuli of positive and negative valence (or the anticipation of a positive and negative US). Slow pupil dilations have previously been shown to respond to emotionally arousing stimuli, but in a valence-unspecific manner (Bradley et al., 2008), similarly as SCR (Hamm et al., 1993; Lang et al., 1990; Lipp et al., 1994). Henderson, Bradley, and Lang (2014) found that the the pupillary light reflex was inhibited during the viewing of both negative and positive (compared to neutral) emotional pictures, and hence valence-unspecific. The same might be the case for the auditory pupil reflex.

The startle reflex seems to hold a distinct position in comparison to other physiological measures, as it is traditionally reported to be modulated by emotional valence (Bradley et al., 1996; Cuthbert et al., 1996; Lang et al., 1990; Vrana et al., 1988). The valence modulation consists of an enhancement of the startle reflex during motivational states of negative valence and an inhibition of the reflex (expressing in decreased response magnitude) during motivational states of positive valence. This reflex may differ from other physiological responses, as it is suggested to be a defensive response to threat (e.g. it causes a protective eye closure reflex in response to loud noises). Bradley and Lang (2007) suggest that the startle reflex during the presence of a CS+ might reflect enhanced 'defense mobilization' in the proximity of threat (the US).

However, the question is whether the fear-potentiation of the startle reflex in conditioning studies relates to the negative valence of CS+, or whether it rather reflects enhanced arousal during US anticipation. Whereas valence refers to the degree of stimulus pleasantness, arousal refers to the degree of excitation (both positive

or negative) elicited by a stimulus (Bradley & Lang, 2007). Both are interconnected in a way that stimuli of high negative or positive valence are judged as more arousing than stimuli of neutral valence. Arousal and valence have therefore shown interactive and additive effects on the startle reflex: the valence modulation of the startle response is strongest for arousing stimuli (Cuthbert et al., 1996) and the valence of picture stimuli seems to contribute additionally to startle response magnitude during concurrent threat of shock (Bublatzky, Guerra, Pastor, Schupp, & Vila, 2013). Lang et al. (1990) accordingly state that arousal can be considered as 'intensity factor' of the startle response.

The valence-modulation of the startle reflex was initially established for probing it during the presence of emotional material (Bradley, Cuthbert, & Lang, 1990). However, in fear conditioning the startle reflex is elicited during the presence of cues which can indicate a following US of negative valence, whereas the cues themselves are often neutral before conditioning. This anticipation (instead of perception) of a US might rather be associated with anticipatory arousal than with actual negative affect. In fact, Sabatinelli, Bradley, and Lang (2001) as well as Dichter, Tomarken, and Baucom (2002) have shown that startle modulation during the anticipation (as opposed to the actual viewing) of emotional pictures was not valence-specific, but modulated by arousal. Correspondingly, several studies have reported a startle potentiation for pleasant CS+ as opposed to neutral cues during conditioning (Mallan & Lipp, 2007; Mallan, Lipp, & Libera, 2008) and even for CS+ which cued a neutral reaction-time task (Lipp, Siddle, & Dall, 2003). If the startle response during outcome anticipation was modulated by valence instead of arousal, it should be inhibited and not increased in anticipation of a positively valenced CS+. Supporting the notion that anticipatory arousal enhances the startle reflex, it has been demonstrated that startle response magnitude increases with temporal proximity to an expected aversive US, i.e. with increasing anticipatory arousal (Grillon, Ameli, Merikangas, Woods, & Davis, 1993).

Apart from arousal, attentional processes have been shown to influence startle magnitude, with enhanced startle responses if conditioned stimuli are attended, irrespective of their valence (Adam, Mallan, & Lipp, 2009; Lipp, Siddle, & Dall, 1997; Mallan et al., 2008). This is also consistent with our finding that startle responses were enhanced during CS- as opposed to inter-trial intervals, just like auditory pupil responses. The traditional view that startle responses, as opposed to other readouts of the conditioned response, are valence-specific is therefore questionable. Instead, habituation and anticipatory arousal are suggested to processes with the largest influence on to startle response magnitude.

Contingency awareness

In our studies, subjective ratings of US expectancy indicated that subjects were aware of reinforcement contingencies, but we did not manipulate or test for contingency awareness, specifically. It has previously been suggested that stimulus discrimination in SCR is dependent on explicit awareness of the CS-US reinforcement contingencies (Sevenster et al., 2014; Tabbert et al., 2011; Tabbert et al., 2006; Weike et al.,

2007). By contrast, startle fear-potentiation can supposedly occur prior to conscious contingency awareness (Hamm & Vaitl, 1996; Sevenster et al., 2014). These findings have fostered the theory that conditioned SCR and startle responses rely on dissociating systems: one for declarative and one for rather reflexive/ procedural fear learning (Hamm & Vaitl, 1996; Hamm & Weike, 2005; Soeter & Kindt, 2010). Following this theory, SCR should theoretically display more similarity with pupil responses, which we suggest to reflect explicit US expectancy. Still, SCR displayed major differences to pupillary measures by habituating throughout fear acquisition. A more differentiated view will be discussed in further detail in context of the neural correlates of conditioned responses below.

Tovote et al. (2015) state that conditioning-induced plasticity in the lateral amygdala temporally precedes cortical plasticity in fear learning. This is possible due to a rapid pathway which reaches the amygdala from the sensory cortices via the thalamus, thereby bypassing a conscious evaluation of the sensory input by higher cortical areas. The amygdala may therefore drive conditioned fear behaviour prior to contingency awareness. Accordingly, the amygdala has been found to activate in response to threat even when conditioned stimuli were presented outside of conscious awareness in human fMRI studies (Critchley, Mathias, & Dolan, 2002; Knight, Waters, & Bandettini, 2009). Physiological measures that are modulated by amygdala activity can therefore potentially differentiate between danger and safety cues even prior to explicit contingency awareness.

4.2 Neuronal circuitry affecting psychophysiological measures during fear learning

4.2.1 Pupillometry

As stated above, the pupillary dilator and sphincter muscles do not receive direct input from higher cortical regions such as the vmPFC or the dACC. The observed correlation of pupil response magnitude with these regions must therefore be mediated indirectly. This is likely given via projections from higher cortical areas to the brainstem and sympathetic or parasympathetic centers, which then target the pupillary muscles. It was mentioned in the introduction that structures like the hypothalamus or the LC receive input from cortical areas and can in turn elicit a number of changes in the physiological periphery by targeting the autonomous nervous system (Habib et al., 2001; Stratakis & Chrousos, 1995).

The locus coeruleus

The pupil has previously been linked closely to activity of the LC (Alnaes et al., 2014; Murphy et al., 2014; Sterpenich et al., 2006), which may be the major relay station from cortical activity to pupil dilations. While the noradrenergic output of the LC has widespread activating effects on brain activity (Aston-Jones & Cohen, 2005), the LC receives input from a large number of brain regions (111 regions of the Allen Brain Atlas in the rodent brain discovered in a tracing study, Schwarz et al., 2015). Many of those regions are associated with fear learning: the LC maintains reciprocal connections to the amygdala and projects to the thalamus

(Samuels & Szabadi, 2008b). Furthermore, there is evidence from animal studies for reciprocal connections between LC and medial prefrontal regions, like the ACC and orbitofrontal cortex (reviewed by Aston-Jones & Cohen, 2005; Del Cid-Pellitero & Garzon, 2011; Marzo, Totah, Neves, Logothetis, & Eschenko, 2014). These interconnections with regions involved in fear learning make the LC a potential hub for fear-related processes and autonomic responses. Giustino and Maren (2018) review the central role of the LC's noradrenergic signaling for fear acquisition and extinction learning. They suggest that the LC can enhance amygdala output (via increased noradrenergic signaling in BLA, specifically) while simultaneously down-regulating prefrontal function during states of fear (where increased levels of noradrenalin are proposed to impair α1- and β-adrenoceptor dependent mechanisms). During low arousal levels instead, the LC is suggested to promote medial prefrontal activity, which can inhibit amygdala output and fear expression. Giustino and Maren (2018) also state that long-term potentiation in the hippocampus can be regulated by noradrenergic signaling, underlining the role of the LC in promoting fear and extinction memory formation. Tanaka, Yoshida, Emoto, and Ishii (2000) manipulated brainstem noradrenergic signaling in a series of pharmacological manipulations in the rat brain. They indeed found that increased noradrenaline in the hypothalamus, amygdala and locus coeruleus provoked stronger fear expression during stress exposure.

Its central position in fear-related processes and its close link to pupil size suggest the LC as a structure which could mediate between cortical activity and pupil dilations. However, the association of LC activity and pupil dilations is not entirely understood. It is discussed as controversial whether there is a direct connection from the LC to the Erdinger-Westphal nucleus, which in turn projects to the pupil (Nieuwenhuis, De Geus, & Aston-Jones, 2011). Alternative accounts suggest that both the LC and the Erdinger-Westphal nucleus receive common inputs from the nucleus paragigantocellularis, or that the superior colliculus relays the signal between LC and Erdinger-Westphal nucleus (Joshi et al., 2016; Nieuwenhuis et al., 2011; Wang & Munoz, 2015). We did not find a significant association of pupil dilations and LC activity in our MR study. However, our fMRI sequence involved whole-brain coverage, a voxel size of 3.4 x 3.4 x 3.4 mm^3 and subsequent spatial smoothing. These settings were not optimized for discovering LC activity, which is itself a small brainstem structure (the human LC has an average size of 12-17 mm x 2.5 mm, Fernandes, Regala, Correia, & Goncalves-Ferreira, 2012). For reproducing the association of LC activity and pupil diameter, other analysis techniques would have been necessary: for example Murphy et al. (2014) related whole-brain fMRI to a specific LC volume of interest, which was informed by high-resolution structural imaging.

The dACC and the salience network

We found an association of pupil response magnitude with salience network activity and, most robustly, with dACC activity during fear learning. The dACC has been linked to a variety of cognitive functions (reviewed by Devinsky, Morrell, & Vogt, 1995), which is why Heilbronner and Hayden (2016) suggest that the dACC generally integrates task-relevant information (such as error monitoring, conflict or outcome values) for

behavioral adjustment. Consistent with this theory, the dACC is a core region of the salience network (together with bilateral insula, thalamus and limbic regions), which is generally associated with states of alertness and responses to motivationally significant stimuli (Coste & Kleinschmidt, 2016; Sadaghiani & D'Esposito, 2015; Seeley et al., 2007). Correspondingly, several studies have suggested an involvement of the dACC as well as the insular cortex in regulating autonomic responses such as fingertip temperature (Yoshihara et al., 2016), the sympathetic component of heart rate variability (Critchley et al., 2003) or SCR during reward feedback anticipation (Critchley et al., 2001). Critchley (2002) suggests that the anterior cingulate cortex in particular might translate higher cognitive processes (for example related to risk or expectancy) into physiological arousal.

Pupil dilations have previously been linked to ACC activity during resting state measurements in the macaque monkey (Aston-Jones & Cohen, 2005; Ebitz & Platt, 2015; Joshi et al., 2016) as well as in humans (Murphy et al., 2014; Schneider et al., 2016). Furthermore, Critchley, Tang, Glaser, Butterworth, and Dolan (2005) reported that activity in rostral ACC and dACC correlated with the magnitude of pupil responses in a numerical STROOP task. Our group recently found further evidence that pupil dilations are related to dACC and salience network activity during reward anticipation (Schneider, Leuchs, Czisch, Samann, & Spoormaker, 2018). It therefore seems that the association of pupil dilations (or other physiological responses) and dACC activity is not task-specific, but found in various cognitive states. In fear conditioning specifically, the dACC has been suggested to be involved in cognitive threat appraisal (Etkin et al., 2011; Maier et al., 2012; Mechias et al., 2010). This is consistent with our findings that pupil dilations were closely related to both dACC activity and subjective ratings of US expectancy.

Findings from the macaque brain indicate that ACC activity can temporally precede, but also follow pupil dilations (Joshi et al., 2016) and the directionality of this association cannot be determined from our analyses. However, it seems plausible that a cognitive evaluation of CS+ and CS- (involving the dACC) temporally precedes a transfer (for example via the LC) to autonomous arousal and pupil dilations.

The thalamus

We found a decline of tonic pupil diameter in both studies and throughout all task phases. When weighting phasic pupil responses by their tonically declining baseline diameter, we found them to correlate with declining thalamus activity in our MR study. This decline in thalamic activity was also revealed by a linearly declining time modulation to all stimuli. Both the decline in thalamic activity and in tonic pupil diameter might therefore be rather unrelated to the fear learning process and might reflect a general decline in vigilance or arousal throughout the task. The thalamus is located within the ascending reticular activating system (ARAS), which originates in the brainstem and has a wakefulness-promoting function (Aston-Jones & Cohen, 2005). Central medial thalamic regions have previously been found to be at the origin of sleep initiation in the rodent brain (Baker et al., 2014). In line with this, Kaufmann et al. (2006) found that thalamic

and hypothalamic activity decreased during the transition from wakefulness to sleep in humans. Whereas phasic pupil responses were related to salience network activity, the tonic pupil diameter might hence inform about slow changes in tonic thalamic activity during fear learning.

Again, we did not find significant clusters in LC in association with declining pupil diameter or proceeding time. Yet, theoretical considerations suggest that the LC - and its connections to the thalamus and pupil – might contribute to tonic changes in pupil diameter. Whereas discrete, phasic pupil responses have previously been related to phasic bursts of LC activity, tonic LC firing is thought to inform about the overall state of wakefulness or arousal. Accordingly, tonic LC firing has previously been related to slow, tonic changes in pupil diameter (Aston-Jones & Cohen, 2005; Aston-Jones, Rajkowski, & Cohen, 1999). Furthermore, a negative association of phasic and tonic firing states of the LC has previously been reported (Aston-Jones & Cohen, 2005; Aston-Jones et al., 1999). This is similar to the inverse relationship of phasic and tonic pupil dilations and further supports the notion that the LC may modulate phasic as well as tonic changes in pupil diameter. Simultaneous pupillometry /fMRI measurements with parameters optimized to detect LC activity can help clarifying this question.

The amygdala

Even though we found activity in right amygdala to be associated with the CS+ > CS- stimulus contrast, pupil responses were not significantly correlated with amygdala activity. As stated in the introduction, animal studies have shown that the amygdala is a structure which is central to the fear learning circuitry (Tovote et al., 2015). However, its activity is inconsistently found in human fear conditioning studies employing functional brain imaging. In our MR study, a linearly declining time modulation to conditioned stimuli revealed that bilateral amygdala activity declined significantly throughout fear acquisition. This has previously been reported in rodent (Brydges et al., 2013; Quirk, Armony, & LeDoux, 1997), as well as in human fear conditioning studies (Büchel, Morris, Dolan, & Friston, 1998; LaBar, Gatenby, Gore, LeDoux, & Phelps, 1998; Lindner et al., 2015; Schneider et al., 1999). This robust decline in amygdala activity throughout fear acquisition might best explain why human imaging studies do not always detect amygdala activity in the average CS+ > CS- contrast. Whereas the amygdala seems to be involved in initial stages of fear learning, its involvement might decline once a fear memory has been formed. Therefore, the amount of averaged trials (spanning only early fear acquisition or also later task phases) likely determines whether amygdala activity reaches significance in the average CS+ > CS- contrast or not not. Another factor which might hinder the detection of amygdala activity is the low signal to noise ratio in fMRI, meaning that a few initial trials might not suffice for a significant result. These considerations also emphasize the importance of viewing fear conditioning as a dynamic learning process and suggest that average contrasts are not appropriate for capturing ongoing cognitive processes. Similarly, the involvement of rather small cell assemblies, such as the amygdala or the LC, is hard to detect with whole brain coverage. Amygdala activity

might hence be better approximated by amygdala-modulated physiological responses throughout fear learning.

Whereas slow pupil dilations do not seem to primarily reflect amygdala activity, this might be different for other physiological readouts. It is unknown whether the more reflexive pupillary responses to startle or light probes relate more strongly to amygdala activity than slow pupil dilations. However, our findings show that auditory pupil responses do not habituate during fear acquisition, which is also the case for amygdala activity and for other physiological measures. This argues against a strong dependency of auditory pupil responses on amygdala activity.

4.2.2 Neural correlates of other physiological readouts of the conditioned response

Similarly as for pupil dilations, the final relay station between cortical activity and physiological responses is likely the brainstem and structures addressing the sympathetic and parasympathetic branches of the autonomous nervous system. Still, physiological response modulation can be associated with different pathways and brain structures.

The neural circuitry underlying the startle reflex is well explained in the rodent brain (see Davis, 1989, 2006) and this reflex is suggested to be mainly amygdala-modulated. The primary reflexive pathway of the startle response receives input from the central amygdala (Davis, 1989; Rosen, Hitchcock, Sananes, Miserendino, & Davis, 1991), which is proposedly why the startle reflex is enhanced during states of fear (Davis, 2006). This is further supported by studies in the macaque monkey, which have shown that the acquisition (however, not retention) of the fear-potentiated startle reflex is abolished with amygdala lesioning (Antoniadis, Winslow, Davis, & Amaral, 2009; Nagai, Critchley, Featherstone, Trimble, & Dolan, 2004).

While the circuitry of the startle reflex is well-described in animal research, there are currently only few published studies that assessed the acoustic startle reflex with simultaneous fMRI during human fear conditioning (Lindner et al., 2015; Van Well, Visser, Scholte, & Kindt, 2012). This is mostly due to technical difficulties like loud background noise as well as electromagnetic inferences on EMG recordings caused by the fMRI acquisition. Lindner et al. (2015) found that startle responses during CS+ (as opposed to CS- and the inter-trial interval) were associated with activity in anterior insula, ACC, thalamus and periaqueductal grey during early extinction. However, this contrast still contained the between stimulus differences in startle responses (which are generally larger to CS+ than to CS-). The reported activity pattern is therefore inevitably confounded by the CS+ > CS- stimulus contrast. Accordingly, it shows large overlap with the salience network, which is typically associated with the CS+ > CS- contrast. Van Well et al. (2012) demonstrated that the CS+ > CS- difference, determined by startle responses, was also reflected in differences in amygdala activity by splitting subjects into a discriminating and non-discriminating group. Neither study reported trial-wise correlations of BOLD signal with startle response magnitude. Therefore, direct evidence

for the association of startle response magnitude and amygdala activity is still missing in human imaging studies.

Similarly, the neural correlates of SCR during fear learning are not fully explored. Even though numerous studies have simultaneously assessed fMRI and SCR during fear conditioning (e.g. Cacciaglia, Pohlack, Flor, & Nees, 2015; Milad et al., 2007; Spoormaker et al., 2011; Tabbert et al., 2011), few studies have addressed the direct neural correlates of SCR. In a review on the neural basis of SCR, Dawson, Schell, and Filion (2007) suggest that SCR can be influenced by brain activity on different levels, i.e. on the highest level by the cortex (such as premotor and frontal regions), on a lower level by the hypothalamus (in association with indirect inputs from the amygdala) or on the lowest level from brainstem regions such as the reticular formation. The final pathway for the initiation of SCR is suggested to pass via the hypothalamus and the brainstem, which then signal to preganglionic sympathetic neurons that ultimately influence the activity of eccrine sweat glands.

In a review, Critchley (2002) concludes that a number of brain regions have the potential to elicit SCR. These comprise the vmPFC, ACC, insular cortex, amygdala, dorsolateral prefrontal and parietal regions. For fear conditioning studies, there is indirect evidence for an association of SCR and the dACC: Linnman, Zeidan, Pitman, and Milad (2013) found that the resting metabolism in dACC predicted differential SCR (CS+ > CS- discrimination) in a subsequent fear conditioning task. Milad et al. (2007) found that differential SCR were related to dACC activity and thickness in fear conditioning. Further evidence points towards an association of ACC activity and skin conductance in the context of emotional or anticipatory arousal: SCR in response to emotionally arousing pictures and electric stimulation were associated with cingulate and motor cortex activity in a PET study (Critchley et al., 2001; Fredrikson et al., 1998). Additionally, SCR during reward feedback anticipation were related to ACC activity in an fMRI study (Critchley et al., 2001). Despite missing evidence from trial-wise associations of SCR magnitude and dACC activity, these results suggest that, in fear conditioning, SCR relate to threat evaluation occurring in dACC, similar to slow pupil dilations.

Phasic SCR and the tonic skin conductance level (SCL) may furthermore relate to different activity patterns: Nagai et al. (2004) found that activity in ACC, insula, thalamus and hypothalamus (among others) was associated with transient SCR, whereas activity in vmPFC and orbitofrontal cortex was related to tonic SCL, irrespective of the employed task (during biofeedback, arousal and relaxation tasks). Tonic and phasic skin conductance may therefore inform about different neuronal processes during fear learning, similar to pupil dilations.

4.2.3 Differentiated view on the neural correlates of startle responses and SCR

The potentially pre-conscious startle reflex modulation and the association of SCR with contingency awareness have led to the assumption that the neural correlates of these two measures dissociate (Hamm & Vaitl, 1996; Hamm & Weike, 2005; Soeter & Kindt, 2010). The findings discussed above indeed indicate that

startle responses are predominantly amygdala-modulated, whereas SCR have primarily been related to dACC activity during fear conditioning. However, an account of entirely dissociating neural underpinning seems to be oversimplified for several reasons.

First, both measures habituate strongly during fear acquisition. SCR thereby display more similarity with startle responses than with pupil dilations, which suggests that they share at least some common underlying mechanisms. Second (and in line with the previous argument), SCR are often associated with the dACC, but they have also been related to amygdala activity: several imaging studies have shown an association of threat-related SCR with amygdala activity (Williams et al., 2006; Williams et al., 2001; Wood, Ver Hoef, & Knight, 2014) and Cacciaglia et al. (2015) reported that left amygdala volume predicted the magnitude of differential SCR during fear acquisition. Accordingly, it has previously been shown that differential SCR also occur to masked conditioned stimuli (Esteves, Parra, Dimberg, & Ohman, 1994; Lipp, Kempnich, Jee, & Arnold, 2014; Ohman & Soares, 1993) or subliminal threat cues (Williams et al., 2006). This indicates that SCR do not necessarily require conscious awareness of the reinforcement contingencies but that they can also relate to a more reflexive, amygdala-dependent fear circuitry. Whereas amygdala activity does not seem necessary for SCR initiation (as they can still occur after amygdala damage, see Tranel & Damasio, 1989), amygdala activity might still contribute to SCR, maybe particularly in response to emotionally salient stimuli (Critchley, 2002). Third, the amygdala is interconnected with various cortical regions like insular cortex or thalamus (Davis, 2007). Startle response magnitude might thereby not solely relate to amygdala activity but, as suggested above, also to general arousal at the time of probing. This physiological arousal might be determined by regions of the salience network like the dACC, which can in turn interact with the amygdala. For example Yoshihara et al. (2016) demonstrated that the functional connectivity from amygdala to dACC and insula increased during the perception of horror movies and that this association was partly modulated by subjective fear.

It can therefore be concluded that higher cortical areas like the dACC have the potential to modulate amygdala activity (and in turn startle response magnitude), while the amygdala can also influence skin conductance. A complete dissociation of the circuitry underlying startle responses and SCR is hence not given.

4.3 Methodological considerations for readouts of the conditioned response

After the discussion of several cognitive-affective and neural processes influencing pupillometric measures, SCR and startle responses, additional consideration should be given to methodological aspects, such as the scoring of physiological responses. The compared measures differ fundamentally in their response latencies: while startle responses occur commonly within 120 ms after the startle probe (Blumenthal et al., 2005), pupil

responses can be initiated after approximately 200 ms (Sirois & Brisson, 2014), as opposed to SCR, which are initiated earliest 1 s after stimulation (Society for Psychophysiological Research, 2012).

Another important differentiation should be made between continuous and triggered responses. Triggered, reflexive responses (e.g. startle, auditory pupil reflex or pupillary light reflex) occur in a rather stereotypical manner and their scoring is rather unambiguous. By contrast, SCR and slow pupil dilations can be scored in a variety of ways as they can be defined to span a rather arbitrary amount of seconds throughout the stimulus interval. While different phases of fear learning may be dominated by different cognitive processes, the same is likely the case for different temporal windows throughout single stimulus intervals. In the course of a CS presentation, early automatic orienting to the CS might shift to a more cognitive evaluation and finally to anticipatory processes (such as US anticipation) towards the end of the trial. Relatedly, the neural activation likely changes throughout single stimulus intervals. As a consequence, the choice of the outcome window for scoring SCR or slow pupil dilations is decisive for the question which cognitive-affective processes and corresponding neural activity patterns are captured (see Figure 3). The considerable heterogeneity in employed scoring methods of physiological responses can contribute substantially to replicability problems (Lonsdorf et al., 2017) and should be considered with care.

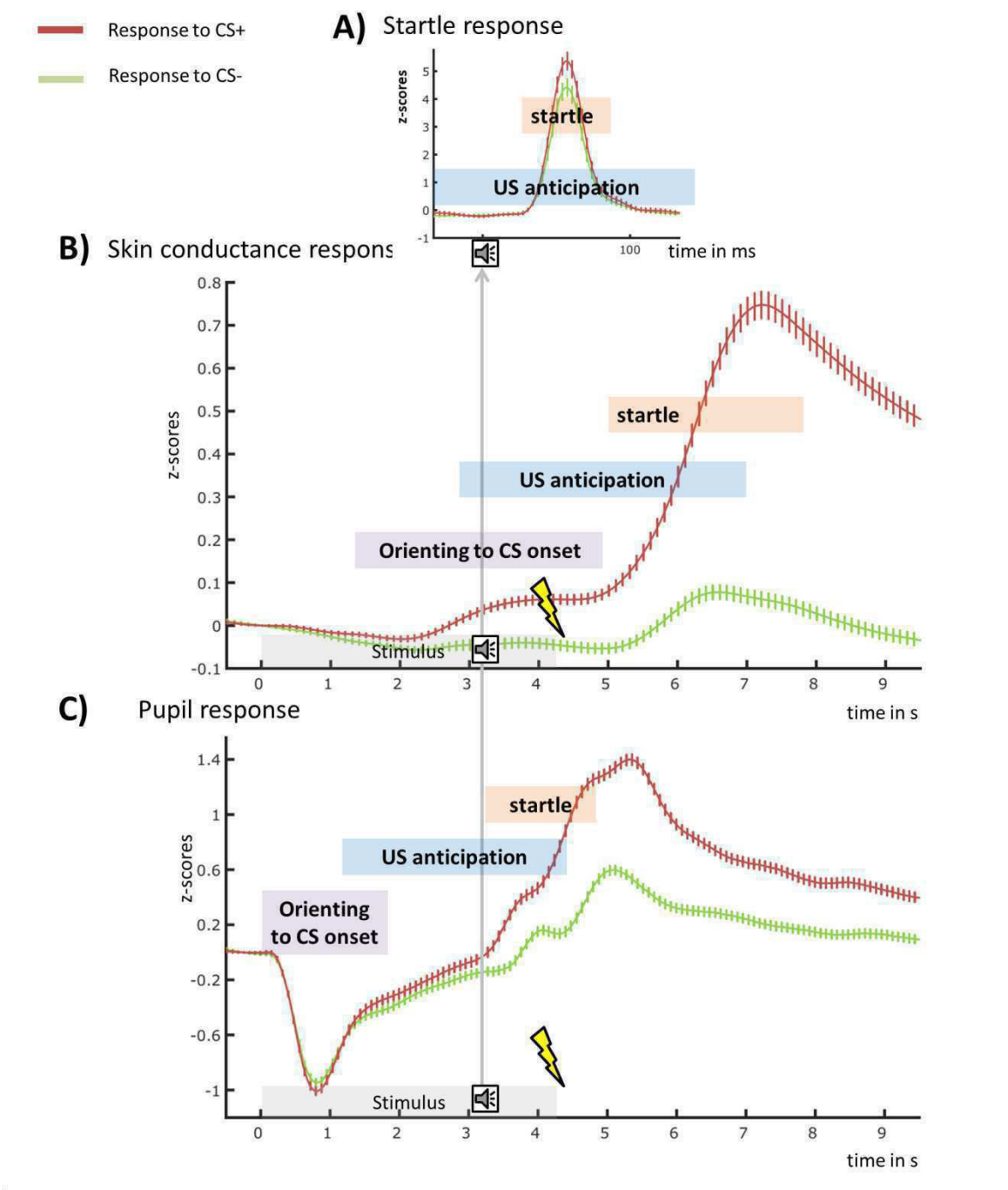


Figure 3. Conditioned response, measured with pupillometry, skin conductance and startle EMG. Responses to CS+ and CS- were averaged across the last block (4 trials per condition) of fear acquisition (N=47). Different measures display largely different response latencies, which is visible in the temporal scales. It is suggested that initial orienting and later threat anticipation affect the physiological measures at different time points throughout (or after) the stimulus interval. The magnitude of triggered responses (caused by the startle probe) is suggested to depend on the ongoing state of autonomous arousal at the time of probing (modulated by US anticipation).

4.3.1 The timing of pupil responses

As stated above, the pupil is influenced by both the parasympathetic and the sympathetic branch of the autonomous nervous system. Rather than acting entirely antagonistic, both branches have been suggested to act in concert, resulting in the observable peripheral changes (Berntson, Cacioppo, Quigley, & Fabro, 1994; Quigley & Berntson, 1990). However, both branches act on different time scales: while parasympathetic (mostly cholinergic) influences have relatively specific target organs and dissipate quickly, the sympathetic (mostly adrenergic) response is widespread and dissipates relatively slowly (Bradley & Lang, 2007). Accordingly, as described above, the pupil response can be split into an early and a later component (Geva et al., 2013; Widmann et al., 2018), which are suggested to be determined by early parasympathetic inhibition and later sympathetic activation, respectively (Widmann et al., 2018). Depending on the temporal definition of the pupillary readout, sympathetic or parasympathetic components will have a stronger influence on the quantified response magnitude. Resulting from the discussion in 4.1.2, later responses (> 1s after stimulus onset, close to US proximity) may have the best chance to capture differences in CS saliency and anticipatory arousal. This is supported by the fact that pupil responses did not differentiate at stimulus onset but showed strongest CS+> CS- differences in close US proximity in our MR study (see also Reinhard & Lachnit, 2002).
A special case is given for pupil responses that are elicited during the stimulus interval. As discussed previously, the magnitude of reflexive pupil dilations is suggested to depend on the ongoing state of physiological arousal at the time of probing. The presence of emotionally significant stimuli is likely to be associated with increased sympathetic activation and thereby enhanced dilator muscle activity. This may increase auditory pupil responses to acoustic probes or, in the case of light probes, antagonize constrictor muscle activity and thereby inhibit the pupillary light reflex during the presence of CS+. We therefore propose that pupillary reflexes mostly depend on the time window during the stimulus interval, at which they are elicited. The later during the stimulus interval, the more likely they reflect anticipatory arousal due to US expectancy, and the more they may resemble slow pupil dilations (which had a similar readout window in our study).

4.3.2 The timing of SCR

SCR mainly relate to sympathetic activity (Dawson et al., 2007) and display the largest latency of the compared measures. Comparable to pupil dilations, SCR at stimulus onset likely reflect different processes from SCR captured in later phases of the stimulus interval. Again we propose that a later response window should reflect US anticipation rather than a reflexive response to CS onset. Pineles et al. (2009) compared the properties of SCR that were scored for different intervals of the conditioned response. They found that onset SCR (0-4 s after CS onset), as well as later SCR (4-9.5 s after CS onset) both discriminated between CS+ and CS-. However, the magnitude of early and late responses correlated only weakly, which suggests that they indeed reflect different cognitive processes. In another study, Luck and Lipp (2016) compared onset, late and

entire stimulus interval SCR and confirmed that these different scoring methods indeed yield different results, suggesting that onset SCR reflect orienting while later SCR reflect anticipation.

This temporal account can also help explaining conflicting opinions whether SCR require contingency awareness and how much they diverge from startle responses in terms of their underlying neuronal circuitry. Here, we propose that in particular onset SCR should be more dependent on amygdala activity and therefore also potentially show a CS+ > CS- difference prior to explicit awareness of reinforcement contingencies. By contrast, later SCR may depend more strongly on contingency awareness and display more similarities with slow pupil dilations by reflecting cognitive stimulus evaluation and threat anticipation.

There are some indications studies supporting this notion: Pineles et al. (2009) reported that early, but not late SCR habituated during fear acquisition, similarly as amygdala activity habituates throughout fear learning (as reported in our MR study and described above). Another indication comes from a fear conditioning study by Spoormaker et al. (2011), employing simultaneous skin conductance and fMRI measurements. In this study, the neural correlates of SCR during fear learning were approximated by a trial-wise parametric modulation with the (overall declining) group mean SCR to individual BOLD time courses. These analyses revealed that onset SCR (0.5-4.5 s after stimulus onset) were correlated with activity in amygdala and thalamus, whereas SCR at stimulus offset (0.5-3.5 s after stimulus offset) yielded ACC activity. Another study by Cheng, Richards, and Helmstetter (2007) found that SCR during the first, but not the second half of the CS interval were associated with amygdala activity. However, stimulus presentations were unusually long in this study (20 s), and early responses comprised all SCR within the first 10 s of CS presentations.

Given their association with amygdala activity and their habituation, onset SCR might display some similarity with reflexive startle responses. By contrast, later SCR may reflect conscious CS evaluation and US anticipation and accordingly relate more strongly to dACC activity and show more similarity to pupil dilations. This is consistent with an account by Ohman (2005) that the automatic detection of threat is reflected in amygdala activity, whereas the later evaluation and sustained threat signaling takes place in higher cognitive areas like the ACC. However, our stimulus presentations were too short to investigate late SCR and we did not acquire skin conductance in our MR study, which is why we could not test this hypothesis.

4.3.2 The timing of startle responses

Finally, the startle response is highly reflexive and short, which is why responses are usually scored in the same temporal range and differences in response latencies are mostly ignored. Grillon, Ameli, Woods, Merikangas, and Davis (1991) found slightly shorter response latencies (in the range of 10-20 ms) during the anticipation of threat as opposed to safety. Concerning stimulus timing, the proximity of the startle probe to the US is likely most decisive for startle response magnitude in fear conditioning. Findings by Grillon et al. (1993) found largest responses in the direct anticipation of threat. As described above, we suggest that startle

response magnitude is mostly dependent on the underlying state of arousal at the time of probing. Differences in probe timing would accordingly lead to more or less influence of anticipatory arousal on startle response magnitude.

5. Conclusions

Pupillometry appears well-suited for fear research due to its non-invasive nature, its sensitivity to emotional content, its rather short response latencies and its high compatibility with imaging techniques. We found that pupil responses can be used as a reliable readout of the conditioned response that can inform about ongoing cognitive-affective processes on a single trial level. Pupil responses most closely related to subjective ratings of US probability: they increased in response to CS+ throughout fear acquisition and reflected reinforcement contingencies during early fear acquisition in our MR study. Later in the task, pupil dilations to a fully reinforced CS+ decreased whereas pupil dilation to the partially reinforced CS+ remained at the same level. This suggests that pupil dilation may capture uncertainty (or a cumulative prediction error over time, as suggested by Koenig et al., 2017) in addition to outcome prediction and threat appraisal.

In both studies, phasic pupil dilations were unaffected by habituation. This stands in opposition to other readouts of the conditioned response like onset SCR or startle responses, which habituated significantly during fear acquisition. Statistical analyses confirmed that pupillary readouts differ substantially from startle responses and SCR, revealing significant differences in their temporal dynamics and only weak correlations on a trial level.

We furthermore found that phasic auditory pupil responses, triggered by acoustic startle probes during CS presentations, can serve as an additional readout of the conditioned response: they differentiated between CS+ and CS- during fear acquisition, yielding larger responses to CS+. As these relatively fast responses can be probed at any time during the fear learning process while being unaffected by habituation, auditory pupil responses constitute a promising readout of conditioned fear which complements slow pupil dilations, startle responses and SCR.

Our imaging study revealed that the magnitude of slow pupil dilations was associated with salience network activity and in particular with activity in dACC on a trial level. We found no significant relationship of pupil dilations with amygdala activity, which is central to the reflexive fear circuitry. This is consistent with the account that phasic pupil responses mainly reflect threat anticipation, US expectancy or uncertainty about the stimulus outcome, which are rather declarative aspects of fear learning.

We furthermore found a significant decline of tonic pupil diameter, which appeared consistently across both studies and all task phases. We propose that the tonic pupil diameter, but not phasic pupil responses, reflects habituation and tracks changes in overall wakefulness. When weighting phasic pupil dilations by their declining baseline diameter, they related to thalamic activity in our imaging study. Given that the thalamus is

a wakefulness promoting structure, this finding supports the account that tonic pupil diameter tracks changes in overall vigilance.

The main focus of this work lies on pupillometry as a measure of fear and on the cognitive-affective processes that can be inferred from changes in pupil diameter during fear learning. Cacioppo, Tassinary, and Berntson (2007, page 12) state: "In its idealized form, a psychophysiological marker is defined as a one-to-one, situation-specific (i.e. context-dependent) relationship […]. The psychophysiological marker relationship assumes only that the occurrence of one (physiological response, parameter of a response, or profile of responses) predicts the occurrence of the other (usually a psychological event) within a given context". Various psychophysiological measures can be employed as markers for fear-related processes. The correct mapping of physiological markers to specific cognitive processes is an important part of advancing psychophysiological research. Our comparison of pupillary measures to established readouts contributes to the understanding of communalities and differences between different readouts of conditioned fear. Our results suggest that, as opposed to slow pupil dilations, onset SCR and startle responses are markers of rather reflexive fear, likely depending on amygdala circuitry. By contrast, it is speculated that later SCR in US anticipation might display more similarity with slow pupil dilations, which we found to be correlated with salience network activity.

As this discussion emphasizes, it is important to consider the temporal dynamics of specific measures, the employed scoring methods and the progress of fear learning upon interpreting physiological responses of conditioned fear. A closer inspection of methodological details in existing fear research may also help consolidating seemingly contradictory findings (for example concerning the awareness dependency of SCR, which seems to count for anticipatory, but not onset responses). Even if such methodological factors are taken into account, a more differentiated view than a one-to-one mapping of readouts and psychological processes seems appropriate. Physiological readouts are likely indicative of multiple cognitive-affective processes, but to different extents: startle responses may for example be preferable for the study of reflexive fear and amygdala circuitry, but they may partly still reflect anticipatory arousal, affected by conscious threat appraisal and dACC activity.

A better linkage of specific psychophysiological markers to psychological processes with discrete neural underpinnings can benefit psychiatric research on fear and anxiety. It can be helpful for tailoring outcome measures to a targeted psychological process and its underlying neuronal circuitry. Specific markers may thereby be indicative of a given psychiatric symptom, whereas others may be less suitable (for example Glover et al., 2011 found PTSD to be associated with startle responses but not SCR). The identification and application of objective physiological markers for psychological processes is in line with the approach promoted by the RDoC initiative. Characterizing individuals with physiological markers (relating to identified neuronal circuitry) can advance psychiatric research to a more objective and biological level. In

fear research, circumscribed biological processes contributing to specific fear-related problems can be identified across the boundaries of existing psychiatric diagnosis categories. This can improve the mechanistic understanding of mental disorders and eventually, help us on the way towards individualized treatment. The well-informed use of psychophysiological markers of fear thereby offers an ideal tool to investigate the biological foundations of fear-related disorders.

References

Adam, A. R., Mallan, K. M., & Lipp, O. V. (2009). The effect of emotional and attentional load on attentional startle modulation. *Int J Psychophysiol, 74*(3), 266-273.

Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *J Vis, 14*(4).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5 ed.). Arlington, VA: American Psychiatric Publishing.

Antoniadis, E. A., Winslow, J. T., Davis, M., & Amaral, D. G. (2009). The nonhuman primate amygdala is necessary for the acquisition but not the retention of fear-potentiated startle. *Biol Psychiatry, 65*(3), 241-248.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci, 28*, 403-450.

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biol Psychiatry, 46*(9), 1309-1320.

Bacigalupo, F., & Luck, S. J. (2017). Event-related potential components as measures of aversive conditioning in humans. *Psychophysiology, 55*(4).

Baker, R., Gent, T. C., Yang, Q., Parker, S., Vyssotski, A. L., Wisden, W., . . . Franks, N. P. (2014). Altered activity in the central medial thalamus precedes changes in the neocortex during transitions into both sleep and propofol anesthesia. *J Neurosci, 34*(40), 13326-13335.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology (2nd ed.)* (pp. 142-162). New York, NY, US: Cambridge University Press.

Berntson, G. G., Cacioppo, J. T., Quigley, K. S., & Fabro, V. T. (1994). Autonomic space and psychophysiological response. *Psychophysiology, 31*(1), 44-61.

Bitsios, P., Szabadi, E., & Bradshaw, C. M. (1996). The inhibition of the pupillary light reflex by the threat of an electric shock: a potential laboratory model of human anxiety. *J Psychopharmacol, 10*(4), 279-287.

Bitsios, P., Szabadi, E., & Bradshaw, C. M. (2002). Relationship of the 'fear-inhibited light reflex' to the level of state/trait anxiety in healthy subjects. *Int J Psychophysiol, 43*(2), 177-184.

Bitsios, P., Szabadi, E., & Bradshaw, C. M. (2004). The fear-inhibited light reflex: importance of the anticipation of an aversive event. *Int J Psychophysiol, 52*(1), 87-95.

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology, 42*(1), 1-15.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034.

Bouton, M. E., & Moody, E. W. (2004). Memory processes in classical conditioning. *Neurosci Biobehav Rev, 28*(7), 663-674.

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1990). Startle reflex modification: emotion or attention? *Psychophysiology, 27*(5), 513-522.

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1996). Lateralized startle probes in the study of emotion. *Psychophysiology, 33*(2), 156-161.

Bradley, M. M., & Lang, P. J. (2007). Emotion and Motivation. In G. Berntson, J. T. Cacioppo, & L. G. Tassinary (Eds.), *Handbook of Psychophysiology* (3 ed., pp. 581-607). Cambridge: Cambridge University Press.

Bradley, M. M., Lang, P. J., & Cuthbert, B. N. (1993). Emotion, novelty, and the startle reflex: Habituation in humans. *Behavioral Neuroscience, 107*(6), 970-980.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607.

Briscione, M. A., Jovanovic, T., & Norrholm, S. D. (2014). Conditioned fear associated phenotypes as robust, translational indices of trauma-, stressor-, and anxiety-related behaviors. *Front Psychiatry, 5*, 88.

Brydges, N. M., Whalley, H. C., Jansen, M. A., Merrifield, G. D., Wood, E. R., Lawrie, S. M., . . . Holmes, M. C. (2013). Imaging conditioned fear circuitry using awake rodent fMRI. *PLoS One, 8*(1), e54197.

Bublatzky, F., Guerra, P. M., Pastor, M. C., Schupp, H. T., & Vila, J. (2013). Additive effects of threat-of-shock and picture valence on startle reflex modulation. *PLoS One, 8*(1), e54003.

Büchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain Systems Mediating Aversive Conditioning: an Event-Related fMRI Study. *Neuron, 20*(5), 947-957.

Cacciaglia, R., Pohlack, S. T., Flor, H., & Nees, F. (2015). Dissociable roles for hippocampal and amygdalar volume in human fear conditioning. *Brain Struct Funct, 220*(5), 2575-2586.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2007). *Handbook of Psychophysiology* (J. T. Cacioppo, L. G. Tassinary, & G. Berntson Eds. 3 ed.). Cambridge: Cambridge University Press.

Careaga, M. B. L., Girardi, C. E. N., & Suchecki, D. (2016). Understanding posttraumatic stress disorder through fear conditioning, extinction and reconsolidation. *Neurosci Biobehav Rev, 71*, 48-57.

Cheng, D. T., Richards, J., & Helmstetter, F. J. (2007). Activity in the human amygdala corresponds to early, rather than late period autonomic responses to a signal for shock. *Learn Mem, 14*(7), 485-490.

Chrousos, G. P., & Gold, P. W. (1992). The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. *JAMA, 267*(9), 1244-1252.

Coste, C. P., & Kleinschmidt, A. (2016). Cingulo-opercular network activity maintains alertness. *Neuroimage, 128*, 264-272.

Craske, M. G., Hermans, D., & Vervliet, B. (2018). State-of-the-art and future directions for extinction as a translational model for fear and anxiety. *Philos Trans R Soc Lond B Biol Sci, 373*(1742).

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *Neuroscientist, 8*(2), 132-142.

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2001). Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron, 29*(2), 537-545.

Critchley, H. D., Mathias, C. J., & Dolan, R. J. (2002). Fear Conditioning in Humans: The Influence of Awareness and Autonomic Arousal on Functional Neuroanatomy. *Neuron, 33*(4), 653-663.

Critchley, H. D., Mathias, C. J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B. K., . . . Dolan, R. J. (2003). Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. *Brain, 126*(Pt 10), 2139-2152.

Critchley, H. D., Tang, J., Glaser, D., Butterworth, B., & Dolan, R. J. (2005). Anterior cingulate activity during error and autonomic response. *Neuroimage, 27*(4), 885-895.

Cuthbert, B. N. (2014). Translating intermediate phenotypes to psychopathology: the NIMH Research Domain Criteria. *Psychophysiology, 51*(12), 1205-1206.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: activation and emotion. *Psychophysiology, 33*(2), 103-111.

Davis, M. (1989). Neural systems involved in fear-potentiated startle. *Ann N Y Acad Sci, 563*, 165-183.

Davis, M. (2006). Neural systems involved in fear and anxiety measured with fear-potentiated startle. *Am Psychol, 61*(8), 741-756.

Davis, M. (2007). Chapter 12 - Neural systems involved in fear and anxiety based on the fear-potentiated startle test A2 - Kesner, Raymond P. In J. L. Martinez (Ed.), *Neurobiology of Learning and Memory (Second Edition)* (pp. 381-425). Burlington: Academic Press.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology (2nd ed.)* (pp. 200-223). New York, NY, US: Cambridge University Press.

De Voogd, L. D., Fernandez, G., & Hermans, E. J. (2016). Disentangling the roles of arousal and amygdala activation in emotional declarative memory. *Soc Cogn Affect Neurosci*.

De Voogd, L. D., Fernández, G., & Hermans, E. J. (2016). Awake reactivation of emotional memory traces through hippocampal–neocortical interactions. *Neuroimage, 134*(1), 563-572.

Del Cid-Pellitero, E., & Garzon, M. (2011). Hypocretin1/OrexinA-containing axons innervate locus coeruleus neurons that project to the Rat medial prefrontal cortex. Implication in the sleep-wakefulness cycle and cortical activation. *Synapse, 65*(9), 843-857.

Devinsky, O., Morrell, M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain, 118 (Pt 1)*, 279-306.

Dichter, G. S., Tomarken, A. J., & Baucom, B. R. (2002). Startle modulation before, during and after exposure to emotional stimuli. *Int J Psychophysiol, 43*(2), 191-196.

Duits, P., Cath, D. C., Heitland, I., & Baas, J. M. (2016). High Current Anxiety Symptoms, But Not a Past Anxiety Disorder Diagnosis, are Associated with Impaired Fear Extinction. *Front Psychol, 7*, 252.

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., . . . Baas, J. M. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. *Depress Anxiety, 32*(4), 239-253.

Ebitz, R. B., & Platt, M. L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron, 85*(3), 628-640.

Esteves, F., Parra, C., Dimberg, U., & Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology, 31*(4), 375-385.

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci, 15*(2), 85-93.

Fernandes, P., Regala, J., Correia, F., & Goncalves-Ferreira, A. J. (2012). The human locus coeruleus 3-D stereotactic anatomy. *Surg Radiol Anat, 34*(10), 879-885.

Flory, J. D., & Yehuda, R. (2015). Comorbidity between post-traumatic stress disorder and major depressive disorder: alternative explanations and treatment considerations. *Dialogues Clin Neurosci, 17*(2), 141-150.

Francati, V., Vermetten, E., & Bremner, J. D. (2007). Functional Neuroimaging studies in posttraumatic stress disorder: reviwe of current methods and findings. *Depress Anxiety, 24*(3), 202-218.

Fredrikson, M., Furmark, T., Olsson, M. T., Fischer, H., Andersson, J., & Langstrom, B. (1998). Functional neuroanatomical correlates of electrodermal activity: a positron emission tomographic study. *Psychophysiology, 35*(2), 179-185.

Fredrikson, M., & Ohman, A. (1979). Cardiovascular and electrodermal responses conditioned to fear-relevant stimuli. *Psychophysiology, 16*(1), 1-7.

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Avila-Parcet, A., & Radua, J. (2015). Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry, 21*(4).

Geller, D. A., McGuire, J. F., Orr, S. P., Pine, D. S., Britton, J. C., Small, B. J., . . . Storch, E. A. (2017). Fear conditioning and extinction in pediatric obsessive-compulsive disorder. *Ann Clin Psychiatry, 29*(1), 17-26.

Geva, R., Zivan, M., Warsha, A., & Olchik, D. (2013). Alerting, orienting or executive attention networks: differential patters of pupil dilations. *Front Behav Neurosci, 7*, 145.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci, 10*(2), 252-269.

Giustino, T. F., & Maren, S. (2018). Noradrenergic Modulation of Fear Conditioning and Extinction. *Front Behav Neurosci, 12*, 43.

Glover, E. M., Phifer, J. E., Crain, D. F., Norrholm, S. D., Davis, M., Bradley, B., . . . Jovanovic, T. (2011). Tools for translational neuroscience: PTSD is associated with heightened fear responses using acoustic startle but not skin conductance measures. *Depress Anxiety, 28*(12), 1058-1066.

Greco, J. A., & Liberzon, I. (2016). Neuroimaging of Fear-Associated Learning. *Neuropsychopharmacology, 41*(1), 320-334.

Grillon, C., Ameli, R., Merikangas, K., Woods, S. W., & Davis, M. (1993). Measuring the time course of anticipatory anxiety using the fear-potentiated startle reflex. *Psychophysiology, 30*(4), 340-346.

Grillon, C., Ameli, R., Woods, S. W., Merikangas, K., & Davis, M. (1991). Fear-potentiated startle in humans: effects of anticipatory anxiety on the acoustic blink reflex. *Psychophysiology, 28*(5), 588-595.

Grillon, C., & Baas, J. (2003). A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Clin Neurophysiol, 114*(9), 1557-1579.

Groves, P. M., & Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychol Rev, 77*(5), 419-450.

Habib, K. E., Gold, P. W., & Chrousos, G. P. (2001). Neuroendocrinology of stress. *Endocrinol Metab Clin North Am, 30*(3), 695-728.

Hamm, A. O., Greenwald, M. K., Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1991). The fear potentiated startle effect. Blink reflex modulation as a result of classical aversive conditioning. *Integr Physiol Behav Sci, 26*(2), 119-126.

Hamm, A. O., Greenwald, M. K., Bradley, M. M., & Lang, P. J. (1993). Emotional learning, hedonic change, and the startle probe. *J Abnorm Psychol, 102*(3), 453-465.

Hamm, A. O., & Vaitl, D. (1996). Affective learning: awareness and aversion. *Psychophysiology, 33*(6), 698-710.

Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *Int J Psychophysiol, 57*(1), 5-14.

Heilbronner, S. R., & Hayden, B. Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annu Rev Neurosci, 39*, 149-170.

Henderson, R. R., Bradley, M. M., & Lang, P. J. (2014). Modulation of the initial light reflex during affective picture viewing. *Psychophysiology, 51*(9), 815-818.

Hopkins, L. S., Schultz, D. H., Hannula, D. E., & Helmstetter, F. J. (2015). Eye Movements Index Implicit Memory Expression in Fear Conditioning. *PLoS One, 10*(11), e0141949.

Hourdaki, E., Giakoumaki, S. G., Grinakis, V., Theou, K., Karataraki, M., & Bitsios, P. (2005). Parametric exploration of the fear-inhibited light reflex. *Psychophysiology, 42*(4), 447-455.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry, 167*(7), 748-751.

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *J Cogn Neurosci, 23*(7), 1587-1596.

Joshi, S., Li, Y., Kalwani, Rishi M., & Gold, Joshua I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron, 89*(1), 221-234.

Jovanovic, T., & Ressler, K. J. (2010). How the neurocircuitry and genetics of fear inhibition may inform our understanding of PTSD. *Am J Psychiatry, 167*(6), 648-662.

Kaufmann, C., Wehrle, R., Wetter, T. C., Holsboer, F., Auer, D. P., Pollmacher, T., & Czisch, M. (2006). Brain activation and hypothalamic functional connectivity during human non-rapid eye movement sleep: an EEG/fMRI study. *Brain, 129*(Pt 3), 655-667.

Knight, D. C., Waters, N. S., & Bandettini, P. A. (2009). Neural substrates of explicit and implicit fear memory. *Neuroimage, 45*(1), 208-214.

Kobayashi, N., Yoshino, A., Takahashi, Y., & Nomura, S. (2007). Autonomic arousal in cognitive conflict resolution. *Auton Neurosci, 132*(1-2), 70-75.

Koenig, S., Uengoer, M., & Lachnit, H. (2017). Pupil dilation indicates the coding of past prediction errors: Evidence for attentional learning theory. *Psychophysiology, 55*(4).

Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH Research Domain Criteria Initiative: Background, Issues, and Pragmatics. *Psychophysiology, 53*(3), 286-297.

Krabbe, S., Grundemann, J., & Luthi, A. (2017). Amygdala Inhibitory Circuits Regulate Associative Fear Conditioning. *Biol Psychiatry, 83*(10), 800-809.

LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron, 20*(5), 937-945.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychol Rev, 97*(3), 377-395.

Lang, P. J., McTeague, L. M., & Bradley, M. M. (2016). RDoC, DSM, and the reflex physiology of fear: A biodimensional analysis of the anxiety disorders spectrum. *Psychophysiology, 53*(3), 336-347.

Larsen, R. S., & Waters, J. (2018). Neuromodulatory Correlates of Pupil Dilation. *Frontiers in Neural Circuits, 12*(21).

Lavin, C., San Martin, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Front Behav Neurosci, 7*, 218.

Lindner, K., Neubert, J., Pfannmoller, J., Lotze, M., Hamm, A. O., & Wendt, J. (2015). Fear-potentiated startle processing in humans: Parallel fMRI and orbicularis EMG assessment during cue conditioning and extinction. *Int J Psychophysiol, 98*(3 Pt 2), 535-545.

Linnman, C., Zeidan, M. A., Pitman, R. K., & Milad, M. R. (2013). Reprint of: resting cerebral metabolism correlates with skin conductance and functional brain activation during fear conditioning. *Biol Psychol, 92*(1), 26-35.

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.

Lipp, O. V., Sheridan, J., & Siddle, D. A. (1994). Human blink startle during aversive and nonaversive Pavlovian conditioning. *J Exp Psychol Anim Behav Process, 20*(4), 380-389.

Lipp, O. V., Siddle, D. A., & Dall, P. J. (1997). The effect of emotional and attentional processes on blink startle modulation and on electrodermal responses. *Psychophysiology, 34*(3), 340-347.

Lipp, O. V., Siddle, D. A., & Dall, P. J. (1998). Effects of stimulus modality and task condition on blink startle modification and on electrodermal responses. *Psychophysiology, 35*(4), 452-461.

Lipp, O. V., Siddle, D. A. T., & Dall, P. J. (2003). The effects of unconditional stimulus valence and conditioning paradigm on verbal, skeleto-motor, and autonomic indices of human Pavlovian conditioning. *Learning and Motivation, 34*(1), 32-51.

Loewenfeld, I. E. (1958). Mechanisms of reflex dilatation of the pupil; historical review and experimental analysis. *Doc Ophthalmol Proc Ser, 12*, 185-448.

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., . . . Merz, C. J. (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.

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans - Biological, experiential, temperamental factors, and methodological pitfalls. *Neuroscience & Biobehavioral Reviews, 80*(Supplement C), 703-728.

Lowenstein, O., Feinberg, R., & Loewenfeld, I. E. (1963). Pupillary Movements during Acute and Chronic Fatigue - a New Test for the Objective Evaluation of Tiredness. *Investigative Ophthalmology, 2*(2), 138-157.

Luck, C. C., & Lipp, O. V. (2016). When orienting and anticipation dissociate--a case for scoring electrodermal responses in multiple latency windows in studies of human fear conditioning. *Int J Psychophysiol, 100*, 36-43.

MacDonald, B., & Barry, R. J. (2017). Significance and Novelty effects in single-trial ERP components and autonomic responses. *Int J Psychophysiol, 117*, 48-64.

Maier, S., Szalkowski, A., Kamphausen, S., Perlov, E., Feige, B., Blechert, J., . . . Tuscher, O. (2012). Clarifying the role of the rostral dmPFC/dACC in fear/anxiety: learning, appraisal or expression? *PLoS One, 7*(11), e50120.

Mallan, K. M., & Lipp, O. V. (2007). Does emotion modulate the blink reflex in human conditioning? Startle potentiation during pleasant and unpleasant cues in the picture-picture paradigm. *Psychophysiology, 44*(5), 737-748.

Mallan, K. M., Lipp, O. V., & Libera, M. (2008). Affect, attention, or anticipatory arousal? Human blink startle modulation in forward and backward affective conditioning. *Int J Psychophysiol, 69*(1), 9-17.

Marois, A., Labonte, K., Parent, M., & Vachon, F. (2017). Eyes have ears: Indexing the orienting response to sound using pupillometry. *Int J Psychophysiol, 123*, 152-162.

Marzo, A., Totah, N. K., Neves, R. M., Logothetis, N. K., & Eschenko, O. (2014). Unilateral electrical stimulation of rat locus coeruleus elicits bilateral response of norepinephrine neurons and sustained activation of medial prefrontal cortex. *J Neurophysiol, 111*(12), 2570-2588.

Mechias, M. L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: implications for conscious appraisal of threat. *Neuroimage, 49*(2), 1760-1768.

Michael, T., Blechert, J., Vriends, N., Margraf, J., & Wilhelm, F. H. (2007). Fear conditioning in panic disorder: Enhanced resistance to extinction. *J Abnorm Psychol, 116*(3), 612-617.

Milad, M. R., Furtak, S. C., Greenberg, J. L., Keshaviah, A., Im, J. J., Falkenstein, M. J., . . . Wilhelm, S. (2013). Deficits in conditioned fear extinction in obsessive-compulsive disorder and neurobiological changes in the fear circuit. *JAMA Psychiatry, 70*(6), 608-618.

Milad, M. R., Orr, S. P., Lasko, N. B., Chang, Y., Rauch, S. L., & Pitman, R. K. (2008). Presence and acquired origin of reduced recall for fear extinction in PTSD: results of a twin study. *J Psychiatr Res, 42*(7), 515-520.

Milad, M. R., Pitman, R. K., Ellis, C. B., Gold, A. L., Shin, L. M., Lasko, N. B., . . . Rauch, S. L. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biol Psychiatry, 66*(12), 1075-1082.

Milad, M. R., & Quirk, G. J. (2012). Fear extinction as a model for translational neuroscience: ten years of progress. *Annu Rev Psychol, 63*, 129-151.

Milad, M. R., Quirk, G. J., Pitman, R. K., Orr, S. P., Fischl, B., & Rauch, S. L. (2007). A role for the human dorsal anterior cingulate cortex in fear expression. *Biol Psychiatry, 62*(10), 1191-1194.

Morris, S. E., & Cuthbert, B. N. (2012). Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin Neurosci, 14*(1), 29-37.

Morriss, J., Christakou, A., & van Reekum, C. M. (2015). Intolerance of uncertainty predicts fear extinction in amygdala-ventromedial prefrontal cortical circuitry. *Biol Mood Anxiety Disord, 5*, 4.

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum Brain Mapp, 35*(8), 4140-4154.

Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. R., & Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: a physiological account of a "default mode" of brain function. *Neuroimage, 22*(1), 243-251.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience, 15*(7), 1040-1046.

Nees, F., Heinrich, A., & Flor, H. (2015). A mechanism-oriented approach to psychopathology: The role of Pavlovian conditioning. *Int J Psychophysiol, 98*(2 Pt 2), 351-364.

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology, 48*(2), 162-175.

Norrholm, S. D., Anderson, K. M., Olin, I. W., Jovanovic, T., Kwon, C., Warren, V. T., . . . Bradley, B. (2011). Versatility of fear-potentiated startle paradigms for assessing human conditioned fear extinction and return of fear. *Front Behav Neurosci, 5*, 77.

Norrholm, S. D., & Jovanovic, T. (2018). Fear Processing, Psychophysiology, and PTSD. *Harv Rev Psychiatry, 26*(3), 129-141.

Norrholm, S. D., Jovanovic, T., Vervliet, B., Myers, K. M., Davis, M., Rothbaum, B. O., & Duncan, E. J. (2006). Conditioned fear extinction and reinstatement in a human fear-potentiated startle paradigm. *Learn Mem, 13*(6), 681-685.

Ohman, A. (2005). The role of the amygdala in human fear: automatic detection of threat. *Psychoneuroendocrinology, 30*(10), 953-958.

Ohman, A., & Soares, J. J. (1993). On the automatic nature of phobic fear: conditioned electrodermal responses to masked fear-relevant stimuli. *J Abnorm Psychol, 102*(1), 121-132.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *Int. J. Hum.-Comput. Stud., 59*(1-2), 185-198.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev, 87*(6), 532-552.

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.

Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neurosci Biobehav Rev, 88*, 117-140.

Preuschoff, K., Hart, B. M., & Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Front Neurosci, 5*, 115.

Quigley, K. S., & Berntson, G. G. (1990). Autonomic origins of cardiac responses to nonsignal stimuli in the rat. *Behav Neurosci, 104*(5), 751-762.

Quirk, G. J., Armony, J. L., & LeDoux, J. E. (1997). Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron, 19*(3), 613-624.

Reinhard, G., & Lachnit, H. (2002). Differential conditioning of anticipatory pupillary dilation responses in humans. *Biol Psychol, 60*(1), 51-68.

Reinhard, G., Lachnit, H., & Konig, S. (2006). Tracking stimulus processing in Pavlovian pupillary conditioning. *Psychophysiology, 43*(1), 73-83.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rosen, J. B., Hitchcock, J. M., Sananes, C. B., Miserendino, M. J., & Davis, M. (1991). A direct projection from the central nucleus of the amygdala to the acoustic startle pathway: anterograde and retrograde tracing studies. *Behav Neurosci, 105*(6), 817-825.

Rytwinski, N. K., Scur, M. D., Feeny, N. C., & Youngstrom, E. A. (2013). The co-occurrence of major depressive disorder among individuals with posttraumatic stress disorder: a meta-analysis. *J Trauma Stress, 26*(3), 299-309.

Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2001). Affective startle modulation in anticipation and perception. *Psychophysiology, 38*(4), 719-722.

Sadaghiani, S., & D'Esposito, M. (2015). Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. *Cerebral Cortex, 25*(9), 2763-2773.

Samuels, E. R., & Szabadi, E. (2008a). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. *Curr Neuropharmacol, 6*(3), 235-253.

Samuels, E. R., & Szabadi, E. (2008b). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. *Curr Neuropharmacol, 6*(3), 254-285.

Schneider, F., Weiss, U., Kessler, C., Muller-Gartner, H. W., Posse, S., Salloum, J. B., . . . Birbaumer, N. (1999). Subcortical correlates of differential classical conditioning of aversive emotional reactions in social phobia. *Biol Psychiatry, 45*(7), 863-871.

Schneider, M., Hathway, P., Leuchs, L., Samann, P. G., Czisch, M., & Spoormaker, V. I. (2016). Spontaneous pupil dilations during the resting state are associated with activation of the salience network. *Neuroimage, 139*, 189-201.

Schneider, M., Leuchs, L., Czisch, M., Samann, P. G., & Spoormaker, V. I. (2018). Disentangling reward anticipation with simultaneous pupillometry / fMRI. *Neuroimage, 178*, 11-22.

Schwarz, L. A., Miyamichi, K., Gao, X. J., Beier, K. T., Weissbourd, B., DeLoach, K. E., . . . Luo, L. (2015). Viral-genetic tracing of the input-output organization of a central noradrenaline circuit. *Nature, 524*(7563), 88-92.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience, 27*(9), 2349-2356.

Sehlmeyer, C., Schoning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human fear conditioning and extinction in neuroimaging: a systematic review. *PLoS One, 4*(6), e5865.

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.

Sillito, A. M., & Zbrozyna, A. W. (1970). The activity characteristics of the preganglionic pupilloconstrictor neurones. *J Physiol, 211*(3), 767-779.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdiscip Rev Cogn Sci, 5*(6), 679-692.

Smith, N. B., Doran, J. M., Sippel, L. M., & Harpaz-Rotem, I. (2017). Fear extinction and memory reconsolidation as critical components in behavioral treatment for posttraumatic stress disorder and potential augmentation of these processes. *Neurosci Lett, 649*, 170-175.

Snowden, R. J., O'Farrell, K. R., Burley, D., Erichsen, J. T., Newton, N. V., & Gray, N. S. (2016). The pupil's response to affective pictures: Role of image duration, habituation, and viewing mode. *Psychophysiology, 53*(8), 1217-1223.

Society for Psychophysiological Research. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034.

Soeter, M., & Kindt, M. (2010). Dissociating response systems: erasing fear from memory. *Neurobiol Learn Mem, 94*(1), 30-41.

Sokolov, E. N. (1963). Higher nervous functions; the orienting reflex. *Annual review of physiology, 25*, 545-580.

Sokolov, E. N. (1990). The orienting response, and future directions of its development. *Pavlov J Biol Sci, 25*(3), 142-150.

Spoormaker, V. I., Andrade, K. C., Schroter, M. S., Sturm, A., Goya-Maldonado, R., Samann, P. G., & Czisch, M. (2011). The neural correlates of negative prediction error signaling in human fear conditioning. *Neuroimage, 54*(3), 2250-2256.

Steiner, G. Z., & Barry, R. J. (2011). Pupillary responses and event-related potentials as indices of the orienting reflex. *Psychophysiology, 48*(12), 1648-1655.

Sterpenich, V., D'Argembeau, A., Desseilles, M., Balteau, E., Albouy, G., Vandewalle, G., . . . Maquet, P. (2006). The locus ceruleus is involved in the successful retrieval of emotional memories in humans. *J Neurosci, 26*(28), 7416-7423.

Stratakis, C. A., & Chrousos, G. P. (1995). Neuroendocrinology and pathophysiology of the stress system. *Ann N Y Acad Sci, 771*, 1-18.

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.

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.

Tanaka, M., Yoshida, M., Emoto, H., & Ishii, H. (2000). Noradrenaline systems in the hypothalamus, amygdala and locus coeruleus are involved in the provocation of anxiety: basic studies. *Eur J Pharmacol, 405*(1-3), 397-406.

Tovote, P., Fadok, J. P., & Luthi, A. (2015). Neuronal circuits for fear and anxiety. *Nat Rev Neurosci, 16*(6), 317-331.

Tranel, D., & Damasio, H. (1989). Intact electrodermal skin conductance responses after bilateral amygdala damage. *Neuropsychologia, 27*(4), 381-390.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon Bull Rev*.

Van Well, S., Visser, R. M., Scholte, H. S., & Kindt, M. (2012). Neural substrates of individual differences in human fear learning: evidence from concurrent fMRI, fear-potentiated startle, and US-expectancy data. *Cogn Affect Behav Neurosci, 12*(3), 499-512.

VanElzakker, M. B., Dahlgren, M. K., Davis, F. C., Dubois, S., & Shin, L. M. (2014). From Pavlov to PTSD: the extinction of conditioned fear in rodents, humans, and anxiety disorders. *Neurobiol Learn Mem, 113*, 3-18.

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear Extinction and Relapse: State of the Art. *Annual Review of Clinical Psychology, 9*(1), 215-248.

Visser, R. M., de Haan, M. I., Beemsterboer, T., Haver, P., Kindt, M., & Scholte, H. S. (2016). Quantifying learning-dependent changes in the brain: Single-trial multivoxel pattern analysis requires slow event-related fMRI. *Psychophysiology, 53*(8), 1117-1127.

Visser, R. M., Kunze, A. E., Westhoff, B., Scholte, H. S., & Kindt, M. (2015). Representational similarity analysis offers a preview of the noradrenergic modulation of long-term fear memory at the time of encoding. *Psychoneuroendocrinology, 55*, 8-20.

Visser, R. M., Scholte, H. S., Beemsterboer, T., & Kindt, M. (2013). Neural pattern similarity predicts long-term fear memory. *Nat Neurosci, 16*(4), 388-390.

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: a new measure of emotion? *J Abnorm Psychol, 97*(4), 487-491.

Wallin, B. G. (1981). Sympathetic nerve activity underlying electrodermal and cardiovascular reactions in man. *Psychophysiology, 18*(4), 470-476.

Wang, C. A., Boehnke, S. E., Itti, L., & Munoz, D. P. (2014). Transient pupil response is modulated by contrast-based saliency. *J Neurosci, 34*(2), 408-417.

Wang, C. A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: implications for cognitive modulation of pupil size. *Curr Opin Neurobiol, 33*, 134-140.

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.

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of Background Noise and Sentence Complexity on Processing Demands during Sentence Comprehension. *Frontiers in Psychology, 7*, 345.

Wetzel, N., Buttelmann, D., Schieler, A., & Widmann, A. (2016). Infant and adult pupil dilation in response to unexpected sounds. *Dev Psychobiol, 58*(3), 382-392.

Widmann, A., Schroger, E., & Wetzel, N. (2018). Emotion lies in the eye of the listener: Emotional arousal to novel sounds is reflected in the sympathetic contribution to the pupil dilation response and the P3. *Biol Psychol, 133*, 10-17.

Wilhelm, H., Ludtke, H., & Wilhelm, B. (1998). Pupillographic sleepiness testing in hypersomniacs and normals. *Graefes Archive for Clinical and Experimental Ophthalmology, 236*(10), 725-729.

Williams, L. M., Liddell, B. J., Kemp, A. H., Bryant, R. A., Meares, R. A., Peduto, A. S., & Gordon, E. (2006). Amygdala-prefrontal dissociation of subliminal and supraliminal fear. *Hum Brain Mapp, 27*(8), 652-661.

Williams, L. M., Phillips, M. L., Brammer, M. J., Skerrett, D., Lagopoulos, J., Rennie, C., . . . Gordon, E. (2001). Arousal dissociates amygdala and hippocampal fear responses: evidence from simultaneous fMRI and skin conductance recording. *Neuroimage, 14*(5), 1070-1079.

Wood, K. H., Ver Hoef, L. W., & Knight, D. C. (2014). The amygdala mediates the emotional modulation of threat-elicited skin conductance response. *Emotion, 14*(4), 693-700.

World Health Organization (WHO). (1992). *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)*. Geneva: World Health Organization.

Yoshihara, K., Tanabe, H. C., Kawamichi, H., Koike, T., Yamazaki, M., Sudo, N., & Sadato, N. (2016). Neural correlates of fear-induced sympathetic response associated with the peripheral temperature change rate. *Neuroimage, 134*, 522-531.

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage, 101*, 76-86.

Zeng, Q., Jia, Y., Wang, Y., Zhang, J., Liu, C., & Zheng, X. (2015). Retrospective reversal of extinction of conditioned fear by instruction. *Conscious Cogn, 35*, 171-177.

Acknowledgements

First of all I would like to express my deep gratitude to Victor Spoormaker. You were not only my boss and supervisor; you were a teacher, a mentor in important life decisions, a friend and one of the major reasons why this PhD project was so much fun. You understood how crucial it was for me to study Neurophilosophy in Venice and you bravely risked that Canadian bears might eat me during my productive phase. Thank you for unlimited amounts of time, for encouragement and trust in my abilities, for always supporting me and being on my side no matter who or what I was struggling with (I found the e-mail again). I have learned so much from you and I will miss you incredibly.

Further thanks go to the other members of my thesis advisory committee: Kathrin Koch, thank you for the friendly and encouraging feedback and the inspiring scientific input. Further thanks go to Elisabeth Binder for always having an open ear and for encouraging and enabling me to go on numerous lab visits, conferences and workshops, which let me grow in so many ways. Patrick Khader, thank you for not only supporting me in my thesis committee, but for being an inspiration and a mentor since my master studies. You showed me how fascinating neurocognitive research methods can be and you were right there defending me when I badly needed it.

Michael Czisch, thank you for supporting me ever since I set my foot into the institute. Your appreciation of my work was extremely encouraging throughout all the years and your advice helped me a lot. You are a major reason why this research group always felt like a family and why work was such a nice place to be.

I furthermore owe thanks to the technical assistants Rosa Schirmer, Anna Hetzel and René Schranzer, who were always supportive - and above all huge thanks to Ines Eidner for her decisive help in data acquisition and her lovely, positive presence throughout all the years.

I furthermore want to thank the Graduate School for Systemic Neurosciences for the great education opportunities, travel grants, social activities and the great friends I made in the last years – you know who you are!

Max, thank you for being a best friend, a brother and a partner in crime. It was a pleasure to share all the ups and downs throughout the last years. Thank you for never judging me (except for the ‘uncomfortable extraversion’ part of course) and for sticking around despite knowing me so well. You know we’re not done here (#einmal Georg immer Georg).

Bori and Fu, you have made my last months some of the best of this PhD. Huge thanks also go to the Brai Büro crew Lara, Ben and Pam, Immi, Chris and Doro: thank you all for your encouragement, emotional support, friendship and for just being around.

Finally I want to thank my parents Andrea and Martin. You enabled me to study in the first place and, together with my siblings Simon and Hanna, you have always been an unlimited source of support, strength and love.

List of publications

Leuchs L, Schneider M & Spoormaker VI. Measuring the conditioned response: a comparison of pupillometry, skin conductance and startle electromyography. *Psychophysiology* (accepted for publication on 2nd of August 2018).

Schneider, M., Leuchs, L., Czisch, M., Samann, P. G., & Spoormaker, V. I. (2018). Disentangling reward anticipation with simultaneous pupillometry / fMRI. *Neuroimage, 178*, 11-22.

Leuchs L, Schneider M, Czisch M & Spoormaker VI (2017). Neural correlates of pupil dilation during human fear learning. *Neuroimage, 147*, 186-197.

Fernandez B, Leuchs L, Sämann PG, Czisch M & Spoormaker VI (2017). Multi-echo EPI of human fear conditioning reveals improved BOLD detection in ventromedial prefrontal cortex. *Neuroimage, 156,* 65-77.

Schneider M, Hathway P, Leuchs L & Spoormaker VI (2016). Spontaneous pupil dilations during the resting state are associated with activation of the salience network. *Neuroimage, 139,*189-201.

Leuchs, L. (2014). Der Ohrwurm. Versuch einer kognitionspsychologischen Einordnung. München: GRIN Verlag.

Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation „Tracking fear learning with pupillometry: Psychophysiological and neuroimaging investigations“ selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation „Tracking fear learning with pupillometry: Psychophysiological and neuroimaging investigations“ is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

Munich, 03.12.2018 Laura Leuchs