

# Method development studies for repeatedly measuring anxiolytic drug effects in healthy humans

*Journal of Psychopharmacology*  
24(5) (2010) 657–666  
© The Author(s), 2010.  
Reprints and permissions:  
[http://www.sagepub.co.uk/  
journalsPermissions.nav](http://www.sagepub.co.uk/journalsPermissions.nav)  
ISSN 0269-8811  
10.1177/0269881109103115

F Klumpers *Departments of Experimental Psychology and Psychopharmacology, Utrecht University, The Netherlands.*

JM van Gerven *Centre for Human Drug Research (CHDR), Leiden, The Netherlands.*

EPM Prinssen *CNS Research, Hoffmann-La Roche, Basel, Switzerland.*

I Niklson *Roche, Clinical Development; current address: Novartis, Basel, Switzerland.*

F Roesch *Clinical Research and Exploratory Development, Hoffmann-La Roche, Basel, Switzerland.*

WJ Riedel *Clinical Research and Exploratory Development, Hoffmann-La Roche, Basel, Switzerland.*

JL Kenemans *Departments of Experimental Psychology and Psychopharmacology, Utrecht University, The Netherlands.*

JMP Baas *Departments of Experimental Psychology and Psychopharmacology, Utrecht University, The Netherlands.*

## Abstract

Human experimental models for anxiety may serve as translational tools for translating preclinical psychopharmacological investigations into human studies. For the evaluation of drugs of which pharmacokinetics and pharmacodynamics are unidentified, repeating measurements after drug administration is necessary for characterising the time course of drug effects. In experiment 1, a threat-of-shock paradigm and adaptations of the Trier mental arithmetic test and the Stroop colour naming test were repeated four times within a day to evaluate whether anxiety responses to this test battery remain stable after repeated testing. This procedure was repeated on 4 days in a second experiment to evaluate suitability of the paradigm for a crossover design with multiple sessions. Results indicate no reductions or changes in fear potentiated startle, the main outcome measure for the threat paradigm, over test sessions or days. Skin

conductance responses and subjective ratings under threat-of-shock showed significant fluctuations but also no systematic decline over time. Finally, the threat paradigm and Stroop test resulted in small increases in reported state anxiety while mental arithmetic produced larger effects that diminished after the first test day. It is concluded that especially the startle paradigm could be a useful new instrument for screening new anxiolytic drugs.

## Key words

anxiety; experimental model; fear; skin conductance response; startle reflex

## Introduction

Human experimental models for fear and anxiety have the potential to greatly facilitate the screening for new anxiolytic drugs. Because these models do not rely on patients they offer many practical advantages, most notably in finding study participants and creating more homogeneous study samples. Furthermore, human models are potentially more powerful in terms of predictive validity than animal models. One of the most widely used objective methods for measuring fear and anxiety experimentally is the fear potentiated startle (FPS) paradigm. The startle reflex, generally probed by presentation of a

loud noise, increases in amplitude reliably both in humans, non-human primates and rodents during states of fear and anxiety (Hamm, *et al.*, 1991; Falls, *et al.*, 1997; Walker, *et al.*, 2002; Winslow, *et al.*, 2002; Grillon, 2008). The fact that the FPS-technique is applied in a similar manner in a variety of species, including the use of threat of electric shock as a means of inducing anxiety, provides a considerable advantage in the translation of pre-clinical studies to the human realm (Grillon and Baas, 2003). Animal work using this technique has created a comprehensive overview of the neurobiological and pharmacological systems that are involved in anxiety (Davis, 2006), providing fertile ground for human research to

build on. Recently, research using the FPS played an important role in the early evaluation of a promising anxiolytic by suggesting the use of D-cycloserine as a facilitator of fear extinction (Walker, *et al.*, 2002), which was later validated in human trials (Ressler, *et al.*, 2004; Hofmann, *et al.*, 2006). Finally, fear potentiation of the startle reflex has been found to be increased in patients with anxiety disorders (Grillon, 2008; Grillon, *et al.*, 2008; Lissek, *et al.*, 2008), further adding to its face validity. For these reasons, the FPS is a promising candidate method in the development of novel screening tools to investigate new, potentially anxiolytic compounds.

Validation of different FPS-paradigms with benzodiazepine administration in healthy subjects has been attempted by several laboratories over the past years (Bitsios, *et al.*, 1999; Riba, *et al.*, 2001; Baas, *et al.*, 2002; Graham, *et al.*, 2005; Scaife, *et al.*, 2005; Grillon, *et al.*, 2006). These studies have yielded mixed results (Baas, *et al.*, 2002), possibly reflecting differences in design parameters. One important factor may be the extent to which anxiety is directly linked to a cue (Baas, *et al.*, 2002). In a recent study, benzodiazepines were found to have an effect on contextual fear but not on cued fear (Grillon, *et al.*, 2006). This differentiation is also supported by animal data showing that these two types of fear seem to have distinct neurobiological bases (Walker, *et al.*, 2003). In humans, contextual anxiety has shown the highest sensitivity to benzodiazepines (Bitsios, *et al.*, 1999; Graham, *et al.*, 2005; Grillon, *et al.*, 2006) and was recently found to specifically differentiate between patients with post-traumatic stress disorder and healthy controls (Grillon, *et al.*, 2008). Therefore, the goal of this study was to further develop an FPS paradigm that is specifically designed for measuring anxiolytic drug effects and focuses on the manipulation of contextual fear.

If the FPS human paradigm is to be used in early drug development process, it would be useful to have more flexibility in the method application. Recently, there have been several reports showing that a paradigm that induces FPS by threat-of-shock is repeatable up to 4 days, separated by up to a week (Bitsios, *et al.*, 1999; Baas, *et al.*, 2002; Graham, *et al.*, 2005; Grillon, *et al.*, 2006), making it possible to test multiple doses or substances in a cross-over design. However, testing new molecules with unknown pharmacokinetic (PK) and pharmacodynamic (PD) characteristics involves uncertainty about the timing of tests relative to drug ingestion. This could be resolved by repeating the paradigm at different time points during one day and in this way covering a larger part of the PK and PD curves. Importantly, this approach demands a paradigm that can be repeated without alterations in the anxiety levels that are induced. To the authors' knowledge, multiple repetitions of a threat-of-shock experiment within a day have not been reported so far and it is unknown whether in such a procedure the FPS and other outcome measures are affected by the repetitions. Two other stressor tasks, computerised adaptations of the mental arithmetic component of the Trier Social Stress Test (Kirschbaum, *et al.*, 1993) and the video-recorded Stroop colour naming task (Leite, *et al.*, 1999; Teixeira-Silva, *et al.*, 2004), were added to the threat-of-shock paradigm to create

an anxiety battery that measures different anxiety facets. This study set out to test whether four repetitions of this battery within a day could be performed without reductions in induced anxiety. Subsequently, this procedure was repeated on 4 days to explore whether it would be feasible to use this method for investigating different doses or substances in a cross-over design.

## Methods

The study protocol was approved by the medical-ethical review committee of the University Medical Center Utrecht.

### Subjects

Subjects were recruited by means of flyers distributed on Utrecht University campus and an advertisement on the University website. A total of 10 subjects completed each experiment (study 1: 8 females, study 2: 10 females). All subjects were aged 18–40 and gave written informed consent prior to participation. Average score on the trait portion of the Spielberger state-trait anxiety inventory (STAI; Spielberger, *et al.*, 1970) for these subjects was 33.3 (study 1: 33.4, study 2: 33.2). Subjects satisfied inclusion criteria as per self-report, including no history of hearing problems, colour blindness, cardiac disorders, neurological disease, psychiatric disorders and alcohol or drug dependence. Furthermore, subjects reported to not smoke regularly or use any illicit drugs or psychoactive medication. Two subjects were excluded after inclusion due to a lack of sufficient startle responses during the first test day; another subject was excluded after failure to comply with the instructions for the experiment.

### Procedure

Each subject was first briefly screened by telephone before receiving an invitation for the screening session at the institute. Upon arrival for the screening, the study procedures were explained to the subject, after which they signed the informed consent for the participation in the study. Next, they were seated in the experimental room and filled in a computerised Dutch version of the trait portion of the STAI (Spielberger, *et al.*, 1970; Van der Ploeg and Defares, 1979) and a questionnaire regarding their medical condition and drug use (including tobacco and alcohol). After this, subjects completed a training session for the colour-word Stroop task to familiarise themselves with the task and response mode to reduce practice effects. Next, electrodes were placed for startle EMG recording and subjects underwent a brief startle test to evaluate startle reactivity. Finally, shock electrodes were placed and a shock work-up procedure was performed to individually set the level of shock used during the first day of the study. The work-up consisted of five sample shocks rated by the subject, shock levels were adjusted in order to achieve an intensity that was rated as "quite annoying". When no impediments for

participation were found during the screening, subjects were invited to participate in the experiment.

In study 1, the total experiment consisted of four sessions within one test day. At the start of the test day, subjects were seated in the experimental room and the shock workup procedure was repeated to assure that subjects again rated the previously determined intensity as “quite annoying”. If necessary, the intensity was adjusted. Next, subjects completed a (non-stressful) practice run for the Stroop task and filled in the state anxiety inventory to measure anxiety immediately prior to the test battery. Subsequently, subjects were instructed that during the threat blocks they could receive shocks at any moment and that during safe blocks no shocks would be presented. To reinforce the instructions, before shock blocks shock electrodes were physically attached to the shock generator in front of the subjects and detached in a clearly visible manner prior to each safe block. Following startle habituation, subjects rated their current subjective state of fearfulness (“how afraid do you feel?”), alertness (“how alert do you feel?”) and feelings of unpleasantness (“how unpleasant do you feel?”) using Visual Analogue Scales (VAS). They then were informed about the condition for the next block (safe/threat) by a text on the screen, and the shock electrodes were either attached or detached by the experimenter (depending on the condition of the first block) who thereafter left the test room. After each block, this procedure was repeated, only now subjects also retrospectively rated their subjective state during the preceding block. The interval between blocks was 35 s. After the shock threat experiment, subjects again filled in the state anxiety inventory and received instructions for the Stroop test on the computer screen. Directly after the Stroop, subjects again filled in the state anxiety inventory. After receiving instructions, subjects did 5 min of mental arithmetic and filled in the state anxiety questionnaire to conclude the session. Total duration of each test session was approximately 1 h, not including the shock workup and Stroop practice run, which were only performed once prior to the first session of the test day. The interval between the start of test sessions was kept at 2 h. Between test sessions, subjects rested and were allowed to eat and drink, but not to drink caffeine-containing beverages or smoke. For study 2, the same procedure was followed and now repeated on 4 test days. The interval between consecutive test days was 3–7 days.

### Tasks and stimuli

**Threat-of-shock experiment** The experimental procedure for the shock threat experiment consisted of a habituation phase followed by three safe and three threat blocks of 2.5 min duration each; safe and threat conditions alternated. The habituation phase involved 12 startle probes presented with an interval that was randomly varied between 20 and 30 s. Startle probes were 50-ms bursts of white noise at 106 dB(A) with an instantaneous rise time. The order of threat and safe blocks was balanced across participants, half started with a safe block and the other half with threat. During the safe blocks, an instruction on

the screen read: “SAFE! Shock electrodes are disconnected”. During the threat blocks, the instruction was: “THREAT! Shock electrodes are connected”. Each block contained six startle probes. The interval between probes varied between 20 and 30 s, with a mean interval of 25 s. Shocks were administered at fixed time points for all participants. These were placed according to a preset schedule to reduce predictability and to reinforce the instruction that shocks could really be administered at any time during the threat condition. For example, shocks during the first test day in both studies were administered at the end of the last threat block of test session 2 and at the beginning of the first threat block of test session 4. During the other test days in study 2, shocks were also placed at pre-defined moments (two in test days 2 and 4, and three in test day 3). For electrical stimulation, a train of 150 2-ms pulses was administered at a rate of 200 Hz. Pulse intensities varied between 0.5 and 9 mA as a result of the workup procedure.

**Stroop task** The colour-word Stroop task that was used in this study was a computerised adaptation of the video-recorded anxiogenic Stroop colour-word test (Leite, *et al.*, 1999; Teixeira-Silva, *et al.*, 2004). The task consisted of 20 colour-word trials in experiment 1, this was extended to 40 trials in experiment 2. The trials started with 1 s of fixation (an asterisk presented in the screen centre) followed by one of the colour words “RED”, “GREEN” and “BLUE” (font size 45, type Arial), displayed in an incongruent colour against a white background for 500 ms (e.g. the word “RED” was always displayed in either green or blue letters). Trials were presented in a semi-random order with the restriction that the same word colour was never presented more than twice in a row. Before each run there was a brief refresher practice run of 20 trials. The practice sessions during the screening and at the start of each test day consisted of 60 trials. During the practice, display duration of the colour words and maximum response time were extended to 1 s. Also, there was no auditory feedback and colour words were replaced by “XXXX”.

Subjects were instructed that they had to respond as fast as possible to the colour of the presented words without making any errors. Furthermore, they were told that they were videotaped during the test, and that the videos would be analysed by a team of experts. In fact, although the video camera in the experiment room was indeed installed and on, no recordings were made. Subjects had to respond within 500 ms by pressing the key on a response box that corresponded with the colour in which the letters were displayed. When subjects made an error or responded too late they received a feedback message (“Error!” or “Too late!”) together with a strident sound. This visual feedback stayed on the screen for 1 s after which the next trial began.

**Mental arithmetic task** For the mental arithmetic task, subjects were instructed to perform a series of backward counting calculations at a pace indicated by a recurring 2000-Hz, 160-ms tone that was played every second. Subjects were told that the next answer had to be given before the next beep to conform to

what was defined as the minimal performance criteria (going from a later defined starting number to zero for 4 times in 5 min), a criterion that was actually virtually impossible to satisfy. Subjects were again made to believe that they were recorded by a video camera even though the camera was not actually recording. When subjects were ready to start, the experimenter started the experiment and a starting number (a randomly selected number between 696 and 705) was presented for 5 s with the instruction to count backwards in 7-number steps. Then, "START" was presented and subjects were required to start counting at the set pace. When subjects were too slow, they were instructed by the experimenter to increase their speed. When subjects made a mistake, the experimenter, who was sitting next to the subject with a list of the correct answers, pressed a button and "ERROR!" appeared on the screen for 1 s together with a strident sound and the experimenter repeated the last correct number. After 1 s, "START" was again presented and subjects had to continue counting from the last correct number. After 5 min, "STOP!" was presented and the subject completed the next state anxiety inventory.

### Measurement and apparatus

All physiological recording and amplification was carried out using the Biosemi Active Two system (<http://www.biosemi.nl>) with matching FLAT Active Ag-AgCl electrodes at a sample frequency of 2048 Hz. The startle response was recorded using electromyography (EMG) from the orbicularis oculi muscle under the right eye. Electrodes were placed  $\pm 15$  mm apart. The skin conductance response (SCR) was recorded by two electrodes placed on the thenar and hypothenar eminences of the palm of the left hand. All electrodes were filled with standard electrolyte gel (Signa Gel; Parker Laboratories). Subjective measurements during the threat experiment were performed using computerised VAS ranging from 0 (not at all fearful/unpleasant/alert) to 100 (very fearful/unpleasant/alert). Dutch versions of the Spielberger state and trait anxiety questionnaires were also presented on the computer. Startle probes were delivered through earphones with foam earplugs (Earlink, Aero Company auditory systems, Indianapolis, IN, USA). Shocks were administered through tin cup electrodes on the left wrist, which were connected to a constant current stimulator (Digitimer DS7A, Digitimer Ltd).

### Data reduction and statistical analysis

All psychophysiological data were processed with Vision Analyzer software (BrainProducts). For the startle data, initially the signal was filtered using a 28-Hz, 12-dB/oct high-pass and a 500-Hz, 24-dB/oct low-pass filter (van Boxtel, *et al.*, 1998; Blumenthal, *et al.*, 2005). The data were segmented into epochs, starting 50 ms before onset of the startle probe and ending 200 ms after onset. Finally, the signal was baseline corrected and rectified and a low-pass filter (14 Hz, 24 dB/oct) was applied for smoothing. Startle magnitude was defined as the amplitude of the first peak in the resulting signal within a

25–100 ms latency window. Subsequently, an artefact rejection procedure was performed on the raw, unfiltered data in which trials with baseline (–30 to 20 ms) activity greater than 2 standard deviations from the mean baseline activity were rejected. Null responses were defined as trials in which the standard deviation of the raw signal increased with less than 55% from baseline. Finally, startle data were *z*-transformed based on single-epoch data from all sessions per subject and converted to *T*-scores ( $T = z * 10 + 50$ ). All subsequent analyses were conducted on the *T*-scores.

SCRs were quantified by scoring the peak response between 0.5 s and 5 s after condition onset and computing the difference between the peak response and the mean skin conductance level during a period from 5 to 0 s prior to block onset. Missing data were replaced by the mean of responses to the other two blocks of the same condition for that test session, or when an entire test session was missing by the mean of two nearest test sessions (this occurred only once). Before statistical analysis, SCR data were normalised by taking the natural logarithm ( $\ln(\text{SCR} + 1)$ ).

The first part of the analysis concerned repeating the threat-of-shock paradigm four times within a day (study 1). Because for the first day of study 2 the exact same procedures were followed, these data were also used for this analysis creating a final sample of 20 subjects for study 1. FPS, SCR and VAS data were averaged per condition for each test session and analysed with SPSS 12 for Windows using analyses of variance with repeated measures for startle and SCR magnitudes, and for subjective ratings under the threat and safe conditions. Greenhouse Geisser epsilon corrections for violations of the assumption of sphericity were used if appropriate. To determine whether the difference between threat and safe conditions varied significantly between test sessions, a 2 (threat, safe)  $\times$  4 (test sessions 1–4) repeated measures ANOVA was carried out. For the second experiment, test day was added as a factor creating a 2 (threat, safe)  $\times$  4 (test session)  $\times$  4 (test day) repeated measures ANOVA. In case the repeated measures ANOVA yielded a significant effect, Student's *t*-tests and univariate ANOVAs were used to characterise differences between conditions. For experiment 1, anxiety potentiation during threat was also tested for each test session using *t*-tests on the difference between safe and threat conditions for each test session. Subsidiary analyses consisted of planned *t*-tests to compare differences between the threat and the safe conditions on different test sessions to further confirm results from the ANOVA. For experiment 2, univariate ANOVAs to compare differences in potentiation between test sessions were run for days 2, 3 and 4 separately (day 1 results are reported under experiment 1). Main effects of threat and interactions with test session and test day (for study 2) are always reported, main effects of test day and test session and three-way interactions (threat  $\times$  test session  $\times$  test day) are only reported when significant.

To investigate the stability of FPS across test sessions within subjects, an additional analysis was conducted in which Pearson correlations between mean startle potentiation for test sessions within a day (study 1) and for different test days (study 2) were calculated. All correlations (six for each study)

were then averaged to one average correlation for each study. Because of the relatively small  $N$ , especially in study 2, only averaged correlations are reported. No correlations were calculated for skin conductance and VAS scores because of the limited amount of data that was acquired per test session (three trials per condition) in addition to the small  $N$ .

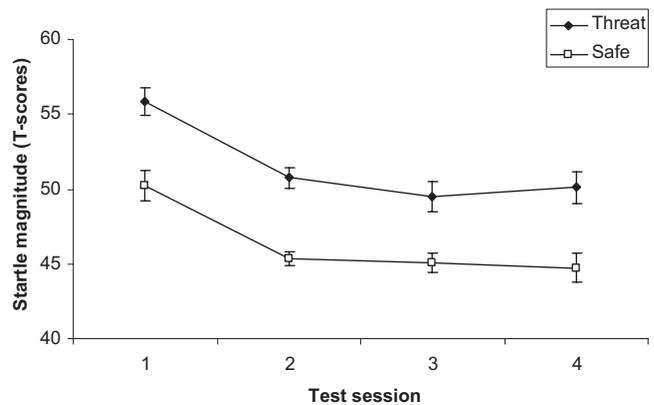
To determine state anxiety fluctuations during the anxiety battery, a repeated measures ANOVA was carried out with time (pre, after FPS, after Stroop, after arithmetic) and test sessions (1–4) as within-subjects factors. For experiment 2, test day (1–4) was again added to the model. Main effects of test day and test session are only reported when significant. In case of a significant time effect, repeated contrasts were used to evaluate effects of each task on state anxiety level by comparing the measurements before and after each test.

## Results experiment 1

### Threat experiment

**Startle** Startle data<sup>a</sup> for threat experiment 1 are displayed in Figure 1. The repeated measurements analysis revealed a highly significant main effect of threat indicating robust startle potentiation during threat blocks ( $F(1,19) = 45.19$ ,  $P < 0.001$ ). Planned  $t$ -tests revealed significant potentiation of the startle response during all test sessions (all  $P$ -values  $< 0.001$ ). Furthermore, a significant main effect of test session indicated that the overall startle response was reduced over consecutive test sessions during the test day ( $F(3,57)$ ;  $Gg\epsilon = 0.73$ ) = 10.83,  $P < 0.001$ ). Importantly, there was no indication for a systematic FPS reduction during the test day as the test session  $\times$  threat interaction turned out to be non-significant ( $F(3,57)$ ;  $Gg\epsilon = 0.47$ )  $< 1$ , n.s.). To further investigate possible differences in FPS between test sessions, exploratory  $t$ -tests were performed to compare threat–safe differences across test sessions. All these tests returned non-significant  $P$ -values, another indication for the absence of any trend in startle potentiation across the four measurements (all  $P$ -values  $> 0.15$ ). The average correlation of FPS between test sessions was  $r = 0.55$ .

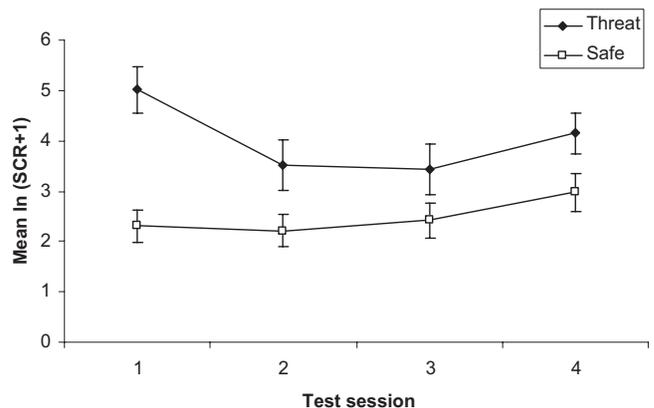
<sup>a</sup> These analyses concern startle amplitude only, latency analyses are not reported. In this study, onset latency reductions under the threat condition were small. The difference between threat conditions only reached marginal significance ( $P = 0.06$ ) over all sessions and was only significant during the first measurement ( $P < 0.05$ ). However, there was no significant test session  $\times$  threat interaction to support the hypothesis that the latency modulation was significantly reduced over sessions ( $P = 0.45$ ). Because the difference between threat conditions was already small during the first measurement ( $\pm 1$  ms) and given that emotional modulation of latencies may be less responsive to anxiolytic drugs than amplitudes (e.g. see Bitsios, *et al.*, 1999 and Graham, *et al.*, 2005), the authors decided not to report the full latency analyses.



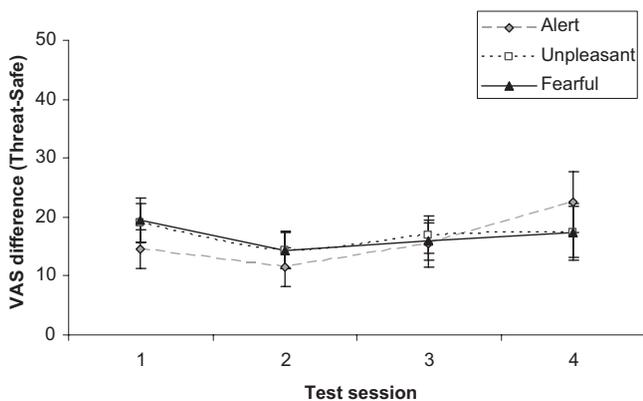
**Figure 1** Mean startle magnitude during threat and safe conditions in the shock threat paradigm for each measurement in experiment 1. Error bars in all graphs represent standard error of the mean.

**Skin conductance** Figure 2 shows the SCRs for the first threat experiment. SCRs were larger in threat vs. safe blocks ( $F(1,19) = 50.09$ ,  $P < 0.001$ ). The planned  $t$ -tests showed significant potentiation of SCR during all test sessions (all  $P$ -values  $< 0.05$ ). Average SCR magnitude differed significantly between test sessions ( $F(3,57)$ ;  $Gg\epsilon = 0.78$ ) = 3.63,  $P < 0.05$ ) and the difference between threat conditions in average SCR also differed between sessions ( $F(3,57)$ ;  $Gg\epsilon = 0.95$ ) = 2.94,  $P < 0.05$ ). SCR potentiation during the first test session was larger when compared to all other test sessions (all  $P$ -values  $< 0.05$ ), while there were no differences between test sessions 2, 3 and 4 ( $P$ -values  $> 0.6$ ).

**Subjective measures** Subjects reported increases in fearfulness, alertness and feelings of unpleasantness during threat blocks when compared to safe blocks ( $F(1,19) = 36.6$ ,  $P < 0.001$ ;  $F(1,19) = 43.3$ ,  $P < 0.001$ ; and  $F(1,19) = 26.8$ ,  $P < 0.001$ , respectively, see Figure 3). Planned  $t$ -tests confirmed



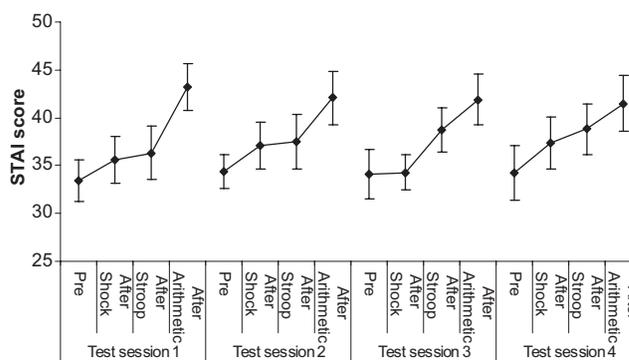
**Figure 2** Mean SCR during threat and safe conditions in the shock threat paradigm for each measurement in experiment 1.



**Figure 3** Mean difference between subjective VAS ratings (ranging from 0 to 100) during threat and safe conditions in the shock threat paradigm for each measurement in experiment 1.

potentiation was present for all test sessions (all  $P$ -values  $< 0.005$ ). Fearfulness and alertness irrespective of threat condition varied over test sessions (fearfulness  $F(3,57; Gg\epsilon = 0.58) = 5.62, P < 0.05$ ; alertness  $F(3,57; Gg\epsilon = 0.71) = 3.25, P < 0.05$ ; unpleasantness  $F(3,57; Gg\epsilon = 0.63) < 1, n.s.$ ). Importantly, however, increases in fearfulness and unpleasantness in threat when compared to safe blocks did not significantly differ between test sessions ( $F(3,57; Gg\epsilon = 0.68) > 1, n.s.$  and  $F(3,57; Gg\epsilon = 0.66) > 1, n.s.$ , respectively). The interaction test session  $\times$  threat reached trend-level significance for alertness ( $F(3,57; Gg\epsilon = 0.72) = 2.56, P = 0.06$ ). This was confirmed by the  $t$ -tests by comparing differences between the threat conditions for each test session. In the case of fearfulness and unpleasantness, no differences between test sessions were observed (all  $P$ -values  $> 0.15$ ). For alertness, there were marginally significant differences between test session 4 and the other test sessions, indicating a larger difference between threat conditions in alertness in the final test session ( $0.05 > P$ -values  $\leq 0.10$ ). While alertness during the safe condition diminished over consecutive test sessions, alertness during the threat blocks of the final test session increased to a level similar as during the first test session (only difference scores plotted in Figure 3).

**Anxiety battery** State anxiety scores significantly increased during the test battery ( $F(3,27; Gg\epsilon = 0.68) = 15.61, P < 0.001$ ; see Figure 4). This increase was not modulated by test session ( $F(9,81; Gg\epsilon = 0.42) = 1.11, P = 0.37$ ). State anxiety significantly increased during the threat-of-shock experiment ( $F(1,9) = 5.38, P < 0.05$ ) and mental arithmetic task ( $F(1,9) = 12.70, P < 0.01$ ) but not significantly so during the Stroop task ( $F(1,9) = 3.82, P = 0.08$ ). Initially, none of these increases seemed significantly modulated by test session as witnessed by non-significant interaction effects ( $F(3,27; Gg\epsilon = 0.76) > 1, n.s.$ ;  $F(3,27; Gg\epsilon = 0.72) = 2.25, P = 0.13$  and  $F(3,27; Gg\epsilon = 0.75) = 1.8, P = 0.19$ , respectively). However, the planned  $t$ -tests indicated that there were significant differences



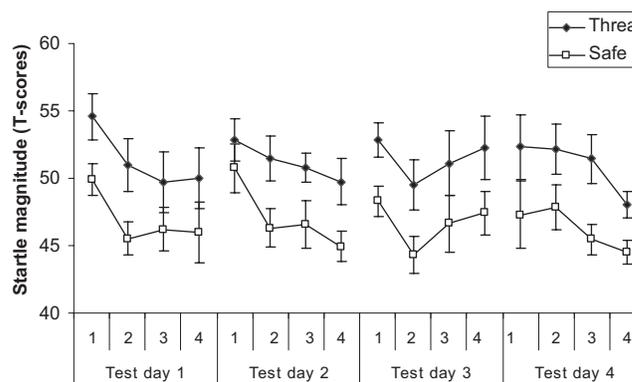
**Figure 4** Mean Spielberger state anxiety scores for each measurement in experiment 1.

between anxiety increases by the Stroop test in test sessions 2 and 4 when compared to test session 3 ( $P$ -values  $< 0.05$ ). Furthermore, state anxiety induced by the Trier was significantly reduced in the third test session when compared to the first measurement ( $P < 0.05$ ).

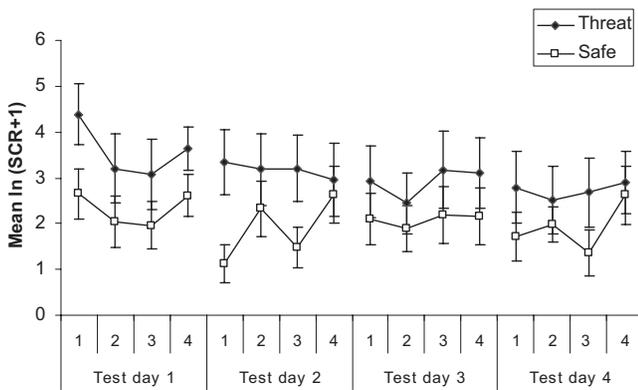
## Results experiment 2

### Threat experiment

**Startle** The startle results for the second threat experiment (see Figure 5) again indicate significant potentiation of the startle response under threat-of-shock ( $F(1,9) = 19.96, P < 0.01$ ). More important, there was no indication of systematic fluctuations in startle potentiation over test days ( $F(3,27; Gg\epsilon = 0.86) < 1, n.s.$ ). Further exploration of potential differences between test sessions indicated that for none of the test days, there were differences in FPS between test sessions (all  $P$ -values  $> 0.15$ ). The average correlation between the mean startle potentiation on different test days was  $r = 0.54$ .



**Figure 5** Mean startle magnitude during threat and safe conditions in the shock threat paradigm for each test session in experiment 2.

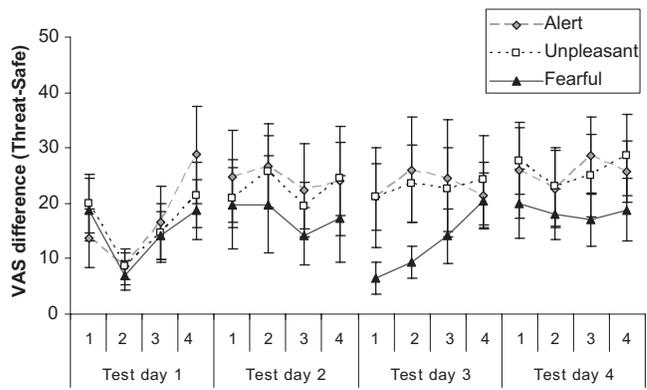


**Figure 6** Mean SCR during threat and safe conditions in the shock threat paradigm for each test session in experiment 2.

**Skin conductance** SCR data for the second threat experiment are illustrated in Figure 6. Subjects responded with larger electrodermal activity to threat vs. safe blocks ( $F(1,9) = 15.99, P < 0.01$ ). This increase did not change systematically over test sessions as evidenced by the non-significance of interactions with test session ( $F(3,27; Gg\epsilon = 0.73) = 1.815, P = 0.187$ ) and test day ( $F(3,27; Gg\epsilon = 0.58) < 1, n.s.$ ). Separate comparisons of the four test sessions on days 2, 3 and 4 reveal a marginally significant test session  $\times$  threat interaction for test day 2 ( $F(3,27; Gg\epsilon = 0.59) = 2.902, P = 0.089$ ; all other test days:  $P > 0.58$ ). On this test day, the increase in SCR during threat blocks was larger during the first test session than during test sessions 2 and 4 ( $P$ -values  $< 0.05$ ).

**Subjective measures** Subjective questionnaire data for experiment 2 as expected indicate clear potentiation of fearfulness, alertness and feelings of unpleasantness during threat vs. safe blocks ( $F(1,9) = 13.45, P < 0.01$ ;  $F(1,9) = 11.93, P < 0.01$ ; and  $F(1,9) = 22.01, P < 0.01$ , respectively, see Figure 7). Increases in arousal and unpleasantness ratings during threat blocks did not significantly change over repeated test sessions ( $F(3,27; Gg\epsilon = 0.45) = 1.22, P = 0.31$ ; and  $F(3,27; Gg\epsilon = 0.56) = 2.07, P = 0.16$ , respectively) and also not over test days (threat  $\times$  test day  $F(3,27; Gg\epsilon = 0.78) = 1.76, P = 0.20$  and  $F(3,27; Gg\epsilon = 0.51) = 1.15, P = 0.33$ , respectively). More elaborate testing of differences between test sessions on test days 2, 3 and 4 again revealed no differences ( $P$ -values  $> 0.3$ ).

Increases in fearfulness appeared to vary somewhat between test sessions (test session  $\times$  threat,  $F(3,27; Gg\epsilon = 0.72) = 3.77, P = 0.04$ ) but not between test days (threat  $\times$  test day,  $F(3,27; Gg\epsilon = 0.74) < 1, n.s.$ ). Post hoc  $t$ -tests indicated that across test days, ratings were significantly lower during test sessions 2 and 3 than during test session 4 ( $P$ -values  $< 0.05$ ). Although the test session  $\times$  threat  $\times$  test day interaction did not reach significance ( $F(9,81; Gg\epsilon = 0.34) = 2.85, P = 0.10$ ), further inspection of the data per test day indicated that during the first test day, there was a significant decrease in fearfulness during the second test session when compared to all other test sessions and that dur-

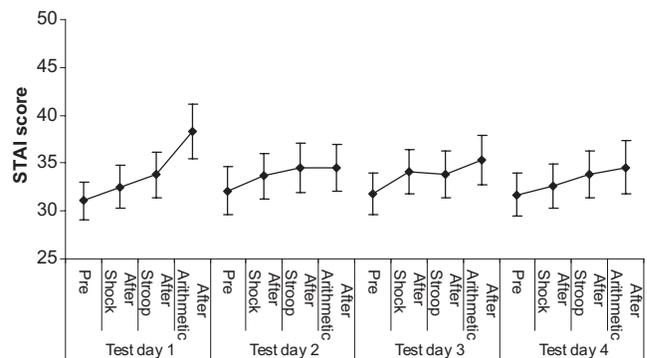


**Figure 7** Mean difference between subjective VAS ratings (ranging from 0 to 100) during threat and safe conditions in the shock threat paradigm for each test session in experiment 2.

ing the third test day ratings during the first test session were lower than during the fourth test session ( $P$ -values  $< 0.05$ ).

**Anxiety battery** State anxiety scores in study 2 were again found to increase during the test battery ( $F(3,27; Gg\epsilon = 0.57) = 8.79, P < 0.01$ ). This increase was not significantly different for different test sessions ( $F(9,81; Gg\epsilon = 0.44) = 1.02, P = 0.41$ ) but differed between test days ( $F(9,81; Gg\epsilon = 0.27) = 3.88, P < 0.05$ ; see Figure 8). In the test days after test day 1, increases in anxiety were less pronounced (see Figure 4) and did not reach significance on the second and fourth day ( $F(3,27; Gg\epsilon = 0.42) = 2.14, P = 0.17$ ;  $F(3,27; Gg\epsilon = 0.92) = 5.63, P < 0.01$  and  $F(3,27; Gg\epsilon = 0.49) = 2.13, P = 0.17$  for test days 2, 3 and 4, respectively).

Looking more specifically at the effects of separate tasks, significant increases after the threat-of-shock experiment ( $F(1,9) = 6.54, P < 0.05$ ) and Stroop task ( $F(1,9) = 9.54, P < 0.05$ ) were found, indicating task-related increases in state anxiety for these tasks but not for the mental arithmetic task ( $F(1,9) = 3.47, P = 0.10$ ). Further analyses showed that only



**Figure 8** Mean state anxiety scores for each test day in experiment 2 collapsed over test sessions.

state anxiety increases by mental arithmetic showed an interaction with test day ( $F(3,27) = 5.83, P < 0.05$ ), caused by a reduction in state anxiety increases by this test after the day 1 (see Figure 8).

## Discussion

This study set out to test a battery developed to evaluate anxiolytic drug effects at multiple time points within a day and across multiple days using FPS as a main readout measure. Results of the first experiment show that when the battery is repeated four times within a day, potentiation of the startle response in the shock threat experiment remains at similar levels while other measures all show some systematic variation. In the second experiment, it was shown that repeating the paradigm four times at four different test days again does not result in significant changes in startle potentiation. Of all measures, the SCR during the threat paradigm and state anxiety increases by mental arithmetic are the only measures that show systematic reductions after repeated measurements.

Previous placebo controlled, cross-over studies have looked at effects of well known anxiolytics at a single time point (Bitsios, *et al.*, 1999; Baas, *et al.*, 2002; Graham, *et al.*, 2005; Grillon, *et al.*, 2006). In this study, we sought to extend the FPS paradigm to potentially allow the measurement of the time course of anxiolytic effects. This may especially be useful when the exact onset of anxiolysis cannot be predicted, as can be the case in early drug development. To the authors' knowledge, no previous studies have been reported that assessed FPS under threat-of-shock on multiple occasions without interventions or substantial changes in the procedure between test sessions. The results of this study indicate that fear potentiation of the startle response does not show systematic habituation in the current paradigm and that therefore consecutive test sessions can be compared validly. In contrast, general startle amplitude (irrespective of condition) was found to habituate during the test day in study 1. Habituation in baseline startle when measurements are repeated is a common observation (e.g. see Quednow, *et al.*, 2006), but at the same time FPS is not reduced (Bradley, *et al.*, 1993; Grillon and Baas, 2002) suggesting that these processes are independent to a certain extent. Habituation in baseline startle calls for the inclusion of an extensive habituation procedure in the threat paradigm to reduce the extra variance caused by the fact that subjects start with either the threat or safe condition.

In contrast to the startle results, in both experiments potentiation of the SCR to the onset of threat conditions relative to safe was found to be reduced after the first measurement. Further inspection of the SCR data showed that the main part of this reduction occurred already during the first block of the first measurement (data not shown). In a previous study that used affective picture stimuli to induce emotion, the habituation of the SCR resulted in a loss in differentiation of responses to emotional and neutral pictures (Bradley, *et al.*, 1993). In the current paradigm, although increases in SCR were reduced

after the first test session, the SCR continued to differentiate between conditions during the four repetitions of this paradigm in study 1 and there were no significant differences in differentiation between the test sessions on each test day in study 2. However, visual inspection of the data indicates that potentiation almost disappeared during the last test sessions of days 2 and 4 supporting the notion that potentiation of the SCR may be more vulnerable to habituation than startle potentiation (Bradley, *et al.*, 1993).

In both study 1 and 2, subjective ratings showed significant variability between test sessions. These variations are likely to be related to the fixed shock reinforcement schedule that was used in this study. The first shock during study 1 and the first test day of study 2 were not administered until the end of the final threat block of the second test session. Subjects may have stopped anticipating the shock after the first test session resulting in the apparent "dip" for all the subjective measures for the second test session. The administration of the second shock (at the start of the first threat block of the fourth test session on day 1) probably resulted in the elevated alertness that subjects reported during threat blocks throughout the fourth test session. Similarly, reductions in fearfulness during the first test session of test day 3 could be caused by the fact that subjects did not receive any shocks during the first test sessions of days 1 and 2 and may not have expected a shock during the first session of day 3. Even though these fluctuations in conscious expectations did not affect the physiological anxiety response level, varying the shock reinforcement schedule per subject is required to eliminate this systematic variation in future studies.

Both the threat-of-shock paradigm and the adapted Stroop task resulted in small increases in reported state anxiety. Furthermore, these effects were not systematically reduced after repetitions of the paradigms within the same test day or on different test days. However, because the paradigms resulted in increases of only a few points on the state anxiety questionnaire that showed some apparently random fluctuations these challenges in their current form seem not especially suited for manipulation with anxiolytic drugs. In previous studies that used the video recorded Stroop in populations with medium trait anxiety, effects were larger (e.g. Leite, *et al.*, 1999; Silva and Leite, 2000; Teixeira-Silva, *et al.*, 2004) probably owing to the fact that these studies included simultaneous exhibition of subject's performance on a screen and a video recorded instruction (Teixeira-Silva, *et al.*, 2004) to make it more plausible that subjects are being monitored by a panel. Also, these studies showed that state anxiety increases are largest when measured during the task instead of after completion of the task. Because other studies that used a digital version of the video-recorded Stroop test also found more substantial effects in a normal population (Hainaut and Bolmont, 2006), it is unlikely that the low increases in state anxiety found in this study are related to the fact that the task was computerised.

The effects on state anxiety of the mental arithmetic task that was employed in this study appeared to be more substantial than those of the other tests and initially are even comparable to those found on in social anxiety patients (Soravia,

*et al.*, 2006) and healthy subjects (Singh, *et al.*, 1999) after completion of the complete Trier Social Stress Test. It, therefore, seems to be a good alternative to the original version that costs less time and personnel. However, the effects diminished (non-significantly) after the first test session and significantly after the first test day. Gradually increasing difficulty of this task may counteract some of these decreases, although none of the subjects reached the preset performance criterion for this task at any test session. For that reason, it is more likely that decreases in experienced anxiety are the result of habituation to the testing procedure. Because an important source of stress in this task (and even more so in the original version) is social-evaluative threat, one may speculate that habituation to the experienced feelings of shame and habituation to the experimenter played an important role. Switching experimenters each test day or even session may be required to keep anxiety responses more stable but can also introduce more variability and is practically demanding.

A general limitation of the measures that were included in this study is that they all have relatively high within-subjects variability when compared to more established parameters for analyses of concentration-effect relationships such as eye movements. Correlations between mean FPS during different test sessions or test days were found to be only in the moderate range (around 0.55 for the *z*-transformed data). This is similar to findings in other studies that correlated startle potentiation during aversive picture viewing (Larson, *et al.*, 2000, 2005), while another study reported lower consistency in responses to emotional pictures (Manber, *et al.*, 2000). Even though multiple repetitions apparently leave the validity of this paradigm intact since startle potentiation is not reduced systematically over time, random variability in startle potentiation within subjects may limit statistical power for detecting drug-related differences between test sessions. On the other hand, the measures used in this study have high face validity for predicting effects on anxiety while more reliable measures primarily give information on sedation. Therefore, incorporation of these measures at an early stage of drug development could potentially make the screening for therapeutic efficacy of anxiolytic compounds more efficient. These exploratory studies suggest that repeated testing using these measures is feasible, although the within-subject variability needs to be addressed further. Although the suitability of repeated FPS-measurements for PK/PD analysis therefore remains to be established, repeated testing offers the additional advantage that it is likely to provide more robust results than a single test at a predefined time point after drug administration.

In conclusion, this study shows that FPS measurements in a threat-of-shock experiment are resistant to systematic habituation effects when repeated within and across days. This indicates that repeated test sessions within a day can be used to characterise drug effects on anxiety at different time points after administration. The other tests in the test battery that was used in this study require further adaptation to increase stability of induced anxiety. Validation of this procedure with a compound with known PK/PD characteristics will be the

next step in further refining the paradigms for screening of new anxiolytic compounds.

### Acknowledgements

The authors would like to acknowledge Anouk Ziech for her help in data acquisition and Nisan Mol for assistance in analysing the data. The study was supported by a restricted grant from Hoffmann-La Roche, Basel, Switzerland.

### References

- Baas, JM, Grillon, C, Bocker, KB, Brack, AA, Morgan 3rd, CA, Kenemans, JL, *et al.* (2002) Benzodiazepines have no effect on fear-potentiated startle in humans. *Psychopharmacology (Berl)* 161: 233–247.
- Bitsios, P, Philpott, A, Langley, RW, Bradshaw, CM, 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: 226–234.
- Blumenthal, TD, Cuthbert, BN, Filion, DL, Hackley, S, Lipp, OV, van Boxtel, A (2005) Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology* 42: 1–15.
- Bradley, MM, Lang, PJ, Cuthbert, BN (1993) Emotion, novelty, and the startle reflex: habituation in humans. *Behav Neurosci* 107: 970–980.
- Davis, M (2006) Neural systems involved in fear and anxiety measured with fear-potentiated startle. *Am Psychol* 61: 741–756.
- Falls, WA, Carlson, S, Turner, JG, Willott, JF (1997) Fear-potentiated startle in two strains of inbred mice. *Behav Neurosci* 111: 855–861.
- Graham, SJ, Scaife, JC, Langley, RW, Bradshaw, CM, Szabadi, E, Xi, L, *et al.* (2005) Effects of lorazepam on fear-potentiated startle responses in man. *J Psychopharmacol* 19: 249–258.
- Grillon, C (2008) Models and mechanisms of anxiety: evidence from startle studies. *Psychopharmacology (Berl)* 199 (13): 421–437.
- 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: 1557–1579.
- Grillon, C, Baas, JM (2002) Comments on the use of the startle reflex in psychopharmacological challenges: impact of baseline startle on measurement of fear-potentiated startle. *Psychopharmacology (Berl)* 164: 236–238.
- Grillon, C, Baas, JM, Pine, DS, Lissek, S, Lawley, M, Ellis, V, *et al.* (2006) The benzodiazepine alprazolam dissociates contextual fear from cued fear in humans as assessed by fear-potentiated startle. *Biol Psychiatry* 60: 760–766.
- Grillon, C, Lissek, S, Rabin, S, McDowell, D, Dvir, S, Pine, DS (2008) Increased anxiety during anticipation of unpredictable but not predictable aversive stimuli as a psychophysiological marker of panic disorder. *Am J Psychiatry* 165: 898–904.
- Hainaut, JP, Bolmont, B (2006) Moderate state-anxiety differently modulates visual and auditory response times in normal- and very low trait-anxiety subjects. *Neurosci Lett* 395: 129–132.
- Hamm, AO, Greenwald, MK, Bradley, MM, Cuthbert, BN, Lang, PJ (1991) The fear potentiated startle effect. Blink reflex modulation as a result of classical aversive conditioning. *Integr Physiol Behav Sci* 26: 119–126.
- Hofmann, SG, Meuret, AE, Smits, JA, Simon, NM, Pollack, MH, Eisenmenger, K, *et al.* (2006) Augmentation of exposure therapy with D-cycloserine for social anxiety disorder. *Arch Gen Psychiatry* 63: 298–304.

- Kirschbaum, C, Pirke, KM, Hellhammer, DH (1993) The 'Trier Social Stress Test'—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28: 76–81.
- Larson, CL, Ruffalo, D, Nietert, JY, Davidson, RJ (2000) Temporal stability of the emotion-modulated startle response. *Psychophysiology* 37: 92–101.
- Larson, CL, Ruffalo, D, Nietert, JY, Davidson, RJ (2005) Stability of emotion-modulated startle during short and long picture presentation. *Psychophysiology* 42: 604–610.
- Leite, JR, Seabra Mde, L, Sartori, VA, Andreatini, R (1999) The video-recorded Stroop Color-Word Test as a new model of experimentally-induced anxiety. *Prog Neuropsychopharmacol Biol Psychiatry* 23: 809–822.
- Lissek, S, Levenson, J, Biggs, AL, Johnson, LL, Ameli, R, Pine, DS, *et al.* (2008) Elevated fear conditioning to socially relevant unconditioned stimuli in social anxiety disorder. *Am J Psychiatry* 165: 124–132.
- Manber, R, Allen, JJ, Burton, K, Kaszniak, AW (2000) Valence-dependent modulation of psychophysiological measures: is there consistency across repeated testing? *Psychophysiology* 37: 683–692.
- Quednow, BB, Kuhn, KU, Beckmann, K, Westheide, J, Maier, W, Wagner, M (2006) Attenuation of the prepulse inhibition of the acoustic startle response within and between sessions. *Biol Psychol* 71: 256–263.
- Ressler, KJ, Rothbaum, BO, Tannenbaum, L, Anderson, P, Graap, K, Zimand, E, *et al.* (2004) Cognitive enhancers as adjuncts to psychotherapy: use of D-cycloserine in phobic individuals to facilitate extinction of fear. *Arch Gen Psychiatry* 61: 1136–1144.
- Riba, J, Rodriguez-Fornells, A, Urbano, G, Morte, A, Antonijoan, R, Barbanj, MJ (2001) Differential effects of alprazolam on the baseline and fear-potentiated startle reflex in humans: a dose-response study. *Psychopharmacology (Berl)* 157: 358–367.
- Scaife, JC, Langley, RW, Bradshaw, CM, Szabadi, E (2005) Diazepam suppresses the acquisition but not the expression of 'fear-potentiation' of the acoustic startle response in man. *J Psychopharmacol* 19: 347–356.
- Silva, FT, Leite, JR (2000) Physiological modifications and increase in state anxiety in volunteers submitted to the Stroop Color-Word Interference Test: A preliminary study. *Physiol Behav* 70: 113–118.
- Singh, A, Petrides, JS, Gold, PW, Chrousos, GP, Deuster, PA (1999) Differential hypothalamic-pituitary-adrenal axis reactivity to psychological and physical stress. *J Clin Endocrinol Metab* 84: 1944–1948.
- Soravia, LM, Heinrichs, M, Aerni, A, Maroni, C, Schelling, G, Ehlert, U, *et al.* (2006) Glucocorticoids reduce phobic fear in humans. *Proc Natl Acad Sci U S A* 103: 5585–5590.
- Spielberger, CD, Gorsuch, RL, Lushene, RE (1970) *Manual for the State-Trait Anxiety Inventory*. Palo Alto, Cal: Consulting Psychologists Press.
- Teixeira-Silva, F, Prado, GB, Ribeiro, LC, Leite, JR (2004) The anxiogenic video-recorded Stroop Color-Word Test: psychological and physiological alterations and effects of diazepam. *Physiol Behav* 82: 215–230.
- van Boxtel, A, Boelhouwer, AJ, Bos, AR (1998) Optimal EMG signal bandwidth and interelectrode distance for the recording of acoustic, electrocutaneous, and photic blink reflexes. *Psychophysiology* 35: 690–697.
- Van der Ploeg, HM, Defares, PB (1979) *Handleiding bij de Zelf-Beoordelings Vragenlijst ZBV: een nederlandstalige bewerking van de Spielberger State-trait Anxiety Inventory STAI-DY*. Lisse: Swets en Zeitlinger.
- Walker, DL, Ressler, KJ, Lu, KT, Davis, M (2002) Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-cycloserine as assessed with fear-potentiated startle in rats. *J Neurosci* 22: 2343–2351.
- Walker, DL, Toufexis, DJ, Davis, M (2003) Role of the bed nucleus of the stria terminalis versus the amygdala in fear, stress, and anxiety. *Eur J Pharmacol* 463: 199–216.
- Winslow, JT, Parr, LA, Davis, M (2002) Acoustic startle, prepulse inhibition, and fear-potentiated startle measured in rhesus monkeys. *Biol Psychiatry* 51: 859–866.