Published in final edited form as:

Psychophysiology. 2019 May ; 56(5): e13325. doi:10.1111/psyp.13325.

Fear-Potentiated Startle Response as an Endophenotype: Evaluating Metrics and Methods for Genetic Applications

Jeanne E. Savage^{1,*}, Ashlee A. Moore¹, Chelsea K. Sawyers¹, Jessica L. Bourdon¹, Brad Verhulst², Dever M. Carney¹, Elizabeth Moroney³, Laura Machlin³, Oumaima Kaabi¹, DR. Scott Vrana⁴, Christian Grillon⁵, Melissa A. Brotman³, Ellen Leibenluft³, Daniel S. Pine³, Roxann Roberson-Nay^{1,6}, and John M. Hettema^{1,6}

¹Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA ²Department of Psychology, Michigan State University ³Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland ⁴Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA ⁵Section on Neurobiology of Fear and Anxiety, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland ⁶Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA

Abstract

The modulation of the startle response (SR) by threatening stimuli (fear-potentiated startle; FPS), is a proposed endophenotype for disorders of the fearful-fearlessness spectrum. FPS has failed to show evidence of heritability, raising concerns. However, the metrics used to index FPS – and, importantly, other conditional phenotypes that are dependent on a baseline – may not be suitable for the approaches used in genetic epidemiology studies. Here we evaluated multiple metrics of FPS in a population-based sample of pre-adolescent twins ($N = 569$ from 320 twin pairs, $M_{age} = 11.4$) who completed a fear-conditioning paradigm with airpuff-elicited SR on two occasions (~1 month apart). We applied univariate and multivariate biometric modeling to estimate the heritability of FPS using several proposed standardization procedures. This was extended with data simulations to evaluate biases in heritability estimates of FPS (and similar metrics) under various scenarios. Consistent with previous studies, results indicated moderate test-retest reliability ($r = .59$) and heritability of the overall SR ($h^2 = 34\%$) but poor reliability and virtually no unique genetic influences on FPS when considering a raw or standardized differential score that removes baseline SR. Simulations demonstrated that the use of differential scores introduces bias in heritability estimates relative to jointly analyzing baseline SR and FPS in a multivariate model. However, strong dependency of FPS on baseline levels make unique genetic influences virtually impossible to detect regardless of methodology. These findings indicate that FPS and other conditional phenotypes may not be well-suited to serve as endophenotypes unless such co-dependency can be disentangled.

*Corresponding author: Jeanne Savage, savagej@vcu.edu. +31205989577.

1. Introduction

The startle response (SR) is an intrinsic reflex that serves to protect the body from potential threats (Lang, 1995). This reflex involves an eyeblink and/or other defensive muscle movements in response to an abrupt stimulus in order to protect vulnerable body parts such as the eyes and neck. Importantly, SR is modulated in response to emotion- and fear-relevant stimuli and is known to be augmented in threatening or negative affective states, with the higher SR during these states relative to neutral states being termed fear-potentiated startle (FPS). FPS is a commonly used measure in psychophysiological research because it is thought to index subjective affective states and is linked to multiple domains of emotional functioning and psychopathology (Grillon & Bass, 2003; Lake, Baskin-Sommers, Li, Curtin, & Newman, 2011; Lang, 1995), highlighting its relevance to the biological systems involved in fear and threat reactivity. Recently, researchers have begun to apply genetic approaches to investigate the etiology of FPS. However, the definition of FPS as a conditional phenotype – an index of change that is dependent on an initial baseline level – may pose problems for the statistical analyses used to carry out genetic research on this measure. We suspect that this issue is not unique to FPS but impacts similar types of research on conditional phenotypes. In the current study, we examine methodological questions in the analysis of FPS, particularly with respect to statistical genetic research.

1.1 Conceptualization of FPS in Research

FPS can be elicited under conditions of contextual threat or aversive emotional arousal, which may be naturally occurring (a dark room; violent images) or learned through association (classical conditioning). Numerous studies have demonstrated that the enhancement of startle in response to a (learned) threat cue co-occurs with cognitive differentiation of the threat cue (i.e. fear learning) and acquired self-reported fear of and autonomic response to the threat cue (e.g. Britton et al., 2013; Glenn, Klein, et al., 2012; Jackson, Payne, Nadel, & Jacobs; 2006; Lau et al., 2008). Furthermore, FPS is directly linked to phobic responses, imminence of threat, and amygdala activation (see Vaidyanathan, Patrick, & Cuthbert, 2009, for a review), indicating that potentiated startle does indeed index affective variation in fearfulness.

A number of different stimuli have been used to elicit the SR, most commonly auditory probes like a blast of white noise or mechanical probes like a puff of air to an area of the skin activating the trigeminal nerve (Blumenthal et al., 2005). As responses to sudden acoustic, tactile, and vestibular stimuli are all integrated as part of a single defensive motivation system (Yeomans, Li, Scott, & Frankland, 2002), these various startle probes produce similar patterns of response with respect to fear/affect potentiation (Blumenthal et al., 2005; Lissek, Baas, et al., 2005; Vaidyanathan et al., 2009), though the overall magnitude of SR may differ across modalities. Furthermore, different types of fear/threat cues like unpleasant pictures, shocks, and screams have been used to modulate startle (Vaidyanathan et al., 2009), though it appears that emotionally valenced stimuli are most relevant to affective systems while threatening (actual or symbolic) stimuli are most relevant to acute fear. Virtually all paradigms have shown an enhancement of SR in the presence of

aversive and/or high-arousal stimuli relative to pleasant/low-arousal stimuli. Studies specifically comparing stimuli have demonstrated FPS similarity between screams and shock (Glenn, Lieberman, & Hajcak, 2012) and between affective images and shock (Lissek et al., 2007), though the overall SR magnitude is greater in the more aversive shock paradigms.

Taken together, the evidence indicates consistency in the phenomenon of modulated startle across multiple domains and modalities. Throughout this manuscript, we use the term FPS to refer to the general phenomenon of SR enhancement in response to aversive or threatening stimuli (though we acknowledge nuances between different designs, particularly between the use of unpleasant vs. specifically threatening stimuli). We focus on FPS because it has been declared a component of the Acute Threat/Fear construct in the Research Domain Criteria (RDoC) classification system (<https://www.nimh.nih.gov/research-priorities/rdoc/constructs/acute-threat-fear.shtml>). Thus, it is a putative intermediate “endophenotype” that should be targeted for genetic research (Cuthbert, 2014). However, the methodological questions we investigate here are broadly generalizable to other types of modulated startle paradigms and, indeed, to other types of non-startle research involving conditional phenotypes.

1.2 Relevance of FPS to Emotional Functioning and Psychopathology

FPS has different clinical correlates at opposite ends of the fearful-fearless continuum (reviewed in Vaidyanathan et al., 2009) and is believed to index sensitivity to fear-relevant cues. Exaggerated FPS has been found in individuals with fear and anxiety disorders (particularly acute fear disorders like phobias), trait fear, and some anxiety-related traits (Lau et al., 2008; Lake et al., 2011; Lissek, Powers et al., 2005; Vaidyanathan et al., 2009), whereas diminished FPS has been found in those with psychopathic traits (Lake et al., 2011; Loomans, Tulen, & van Marle, 2015; Patrick, 1994). Elevated overall SR, but not FPS, has often been found in relation to conditions like generalized anxiety disorder (Vaidyanathan et al., 2009), indicating specificity of FPS to acute threat response. It follows that, by studying individual differences in FPS, dysregulation in brain substrates underlying threat response could be identified.

A key approach to studying sources of individual differences of a trait is the use of genetically-informative study designs. Some aspects of the SR have been found to be heritable; for a thorough review, see Savage et al. (2016). Briefly, studies have found the overall SR (not FPS) to have a heritability of 37% - 67% (Anokhin, Golosheykin, & Heath, 2007; Dhamija, Tuvblad, Dawson, Raine, & Baker, 2017; Vaidyanathan, Malone, Miller, McGue, & Iacono, 2014). These same studies, using negative/positive affective imagery to alter SR, found no genetic contribution to the affective modulation of startle in humans. No heritability studies of SR in the context of a fear-conditioning paradigm have yet been conducted. However, there is evidence that there is a substantial genetic contribution of 46% to FPS in mice (McCaughran, Bell, & Hitzemann, 2000). Additionally, one twin study of fear-conditioning in humans (Hettema et al., 2003) found distinct sets of genetic influences significantly contributing to baseline versus fear-potentiated response. Fear response in that study, however, was measured with skin conductance rather than startle, which may not capture the same components of fear acquisition processes (Hamm & Weike, 2005).

When a physiological trait shows heritability and co-varies with a disorder, it may be considered as a potential “endophenotype” for the disorder (Gottesman & Gould, 2003). Endophenotypes were initially proposed as intermediate biological traits that lie closer to the etiological mechanisms influencing a behavior/disorder, allowing greater insight into its underlying genetic influences. Although more recent investigations have suggested that they may be just as genetically complex as behaviors/disorders (Flint & Munafo, 2007), endophenotypes remain valuable targets because of their position in a mediational pathway that can provide insight into the mechanisms by which genetic variants influence distal psychological outcomes. Despite the mixed findings cited above, it has been postulated that FPS is an endophenotype for anxiety disorders and psychopathy (Patrick, 1994; Savage et al., 2016). To be validated as an endophenotype, FPS must be shown to be (1) associated with the disorder, and (2) heritable, among other secondary criteria (Gottesman & Gould, 2003; see also Cannon & Keller, 2006). FPS clearly meets the first criterion, but studies thus far have shown no evidence for its heritability. Furthermore, from a practical standpoint, endophenotypes should be reliable traits with low measurement error if they are to serve as indices for “noisy” higher-order psychological disorders. One study found good 1-week test-retest reliability of general SR and of startle potentiation in a threat-of-shock task, but there has been relatively low and inconsistent reliability for other, less aversive FPS paradigms like affective picture viewing (Kaye, Bradford, & Curtin, 2016). Longer testing periods are also needed to establish FPS as a stable, trait-like measure of individual differences. If FPS is indeed not a reliable and heritable trait, it has little value for further consideration as an endophenotype.

1.3 Limitations of Current Measures of FPS

A potential limitation of the methodology of existing heritability studies is that FPS has been characterized by subtracting SR during a neutral or positive affect condition (e.g., baseline) from SR during a fear/negative affect condition and estimating heritability based on the raw residuals (Anokhin et al., 2007) or percent change from neutral (Dhamija et al., 2017). Further, some studies have first conducted a within-person standardization of the SR prior to calculating difference scores (e.g. Vaidyanathan et al., 2014). Although there is no consensus in the field, it is common practice to use some sort of standardization. This practice removes variability in the SR due to factors not directly relevant to the potentiation of startle, such as inter-individual differences in magnitude of muscle activity and stochastic variation in experimental conditions (Blumenthal et al., 2005; Grillon & Baas, 2003). This is important when the aim is to link between-individual differences in FPS to between-individual differences in an outcome. However, it is potentially problematic in genetic epidemiology studies because the methods used to estimate heritability in twin samples involve statistical decompositions of that same between-individual variation in the trait of interest into genetic and environmental contributions (Neale & Cardon, 1992). Others have also found that standardized or percent-change scores fail to demonstrate the expected patterns of correlations with external validators (Bradford, Starr, Shackman & Curtin, 2015), suggesting that these transformations may remove important information from FPS measures.

We hypothesize that modifying scores with these differential transformations may be masking true heritability of FPS in the twin-based biometric modeling methodology.

Multivariate twin models that simultaneously include the raw, unstandardized, baseline startle response magnitudes in addition to FPS may provide more accurate estimates of the heritability of both baseline SR and FPS, akin to the statistical improvement achieved by using polynomial regression methods instead of difference scores (e.g. Edwards, 2001). The studies of Anokhin et al. (2007) and Dhamaji et al. (2017) have indirectly (and incidentally) found such an effect through their use of a “common pathway” multivariate model for evaluating genetic and environmental influences of *overall startle*, a composite measure comprised of raw SR values under neutral, positive, and negative affect conditions; however, they treated *affect-modulated startle* as a separate univariate outcome. A systematic investigation into the impact of these measurement issues on biometrical heritability estimates has yet to be conducted. Further, existing studies have focused on adult or adolescent samples (mean age of 15 [Dhamaji et al., 2017], 18 [Vaidyanathan et al., 2014], or range of 18–29 [Anokhin et al., 2007]). Previous work has shown that FPS manifests similarly throughout development from childhood to adulthood, though the magnitude of SR increases from early to late childhood (Quevedo, Smith, Donzella, Schunk, & Gunnar, 2010). It may be that the heritability of FPS differs at earlier ages, particularly in the child-to-adolescent transitional period during which the onset of anxiety symptomology and psychopathic traits often occurs (Lynam et al., 2009; Kessler et al., 2005).

1.4 Aims of the Current Study

This study seeks to address three primary questions surrounding the utility of FPS as an endophenotype for fear-related psychopathology: (1) How reliable are FPS measures over a longer interval? (2) What are the genetic and environmental sources of individual differences in FPS in pre-adolescents? and (3) How do the metrics used to index FPS affect estimates of its heritability? To address these questions, we assess multiple measures of FPS from an airpuff-elicited startle paradigm on two measurement occasions in an epidemiological sample of 9–14 year-old twins, and test a series of univariate and multivariate biometric models. With the broader aim of generalizing our results to other types of modulated startle, and other conditional phenotypes, we conduct simulations that evaluate how data transformations affect the estimates of heritability under multiple scenarios. Given the evolutionary relevance of FPS and the robust associations it demonstrates with fear-related psychopathology, we hypothesize that we will identify genetic factors unique to FPS when baseline startle and FPS are analyzed jointly in a maximally-informative multivariate statistical model. We also predict that the simulation results will demonstrate biases in the heritability estimates derived from standardized or differential scores that help explain limitations in prior approaches. Together, results will elucidate putative endophenotypic properties of FPS and, consequently, our understanding of the biological basis of disorders on the fearful-fearfulness spectrum. Further, results will provide methodological guidance for future (genetic) studies of psychophysiological endophenotypes.

2. Method

2.1 Participants

The current study used data from the Virginia Commonwealth University Juvenile Anxiety Study (VCU-JAS). Twins aged 9–14 were recruited from the Mid-Atlantic Twin Registry

(Lilley & Silberg, 2013) to assess measures that putatively probe psychopathology including self-report and physiological measures collected during laboratory paradigms. Only Caucasian twins were recruited to minimize genetic heterogeneity. Participants were assessed at one of two sites: 1) VCU in Richmond, Virginia, or 2) the National Institute of Mental Health (NIMH) in Bethesda, Maryland. The university's Institutional Review Board approved this study, and all participants and their parent/guardian provided informed consent/assent. For a full description of the study, see Carney et al. (2016).

Of the 796 individuals from 398 complete twin pairs who enrolled in VCU-JAS, 675 participated in the startle paradigm. After data cleaning, including removal of non-responders and outliers (described below), 569 individuals (90 complete monozygotic [MZ] pairs, 159 complete dizygotic [DZ] pairs, and 71 singletons) were available for the current analyses. The final analytic sample was 53.6% female with a mean age of 11.4 ($SD = 1.5$) years. Of these individuals, 481 (84.5%) were assessed at VCU and 88 (15.5%) at the NIMH site. Twins within a pair always completed the laboratory paradigms at the same site and on the same day. Participants were invited back for a second reliability visit (11–59 days apart; $M = 24$ days) in which they were randomized to repeat ~75% of the visit 1 tasks. A total of 274 individuals completed the second visit, of whom 149 individuals (after data cleaning) had complete SR task data at both visits and were included in the reliability analyses here (21 MZ pairs, 40 DZ pairs, and 27 singletons; 53.0% female; $M[SD]$ age 11.5 [1.5]).

2.2 Procedures

2.2.1 Startle paradigm.—The “Screaming Lady” laboratory paradigm was used to assess FPS (for a full description of the paradigm, see Lau et al., 2008 and the supplement of Britton et al., 2013). This task was a differential fear conditioning paradigm in which an aversive loud (95 dB, 500ms), piercing scream (unconditioned stimulus; UCS) was paired with one of two distinct images of female faces. An airpuff was used as the startle-inducing probe, as described below. Previous use of this task has demonstrated its ability to evoke fearful responses by self-reported fear, threat cue learning, and potentiation of startle and skin conductance in response to the UCS and conditioned stimulus (CS+) (Britton et al., 2013; Glenn, Klein, et al., 2012; Jackson et al., 2006; Lau et al., 2008). The use of scream as a UCS evokes a smaller magnitude of SR than more aversive stimuli like shocks but results in a similar potentiation of startle and is more suitable to younger participants (Glenn, Klein, et al., 2012). Also, although the majority of SR research has used auditory startle probes, startle elicitation has been shown to be relatively consistent across different modalities (Blumenthal et al., 2005) and specifically between airpuff and white noise probes, with airpuff resulting in less (extraneous) physiological arousal and being rated as less aversive (Lissek, Baas et al., 2005). Therefore, although this specific task was selected for its suitability for children, it can be considered broadly generalizable to other paradigms using different startle probes and/or threatening stimuli.

During the task, participants sat in a comfortable chair facing a computer screen where the images were displayed, and the room was darkened. Participants wore in-ear headphones through which the UCS was presented. The task involved a 2-minute acclimatization period followed by three successive phases:

1. **Habituation:** 12 startle probes administered in the absence of the UCS and paired with neutral expression images of the two female faces or a blank screen during the inter-trial interval (ITI) (4 presentations of each). This was the baseline measure.
2. **Fear acquisition:** 30 startle probes paired with the neutral expression images of the two faces or during the ITI (10 presentations of each). One face (CS+) was paired with the UCS (scream) on 8 out of the 10 trials to create a measure of fear-potentiated startle. The UCS occurred immediately after presentation of the neutral image, alongside an image of the same face morphed to display a fearful expression. The second face (CS-) was never paired with the UCS.
3. **Fear extinction:** 24 startle probes paired with the neutral expression images of the two faces or a blank screen during the ITI (8 presentations of each; no UCS).

In each phase, images were presented for 8s with startle probes occurring 5–6s after image onset. Blank screens were displayed for a variable ITI period approximately 30s with probes occurring at a variable time 15–20s after onset. Presentation order of startle probes and CS +/CS- designation followed one of four randomization schedules counterbalanced across participants but with the same schedule within twin pairs.

2.2.2 Startle response measure.—Throughout all phases of the task, the SR was repeatedly elicited by the administration of the startle probe, a mechanical airpuff (40ms, 10psi of compressed room air) to the center of the participant's forehead through a polyethylene tube affixed approximately 1cm from the skin via a helmet. Startle probes were automatically administered with a solenoid device connected to the computer running the task using E-Prime software (Psychology Software Tools, Sharpsburg, PA).

Electromyography (EMG) was used to measure the magnitude of the eyeblink SR from the electrical activity of the orbicularis oculi muscle occurring after each startle probe. EMG activity was recorded via two reusable 4 mm Ag/AgCl electrodes filled with a high-conductivity electrode gel and attached via trimmed, double-sided adhesive collars affixed 1cm apart under the participant's left eye. A ground electrode was placed in the center of the participant's forearm. Prior to recording, the participant's skin was prepared by lightly scrubbing with an exfoliant gel (NuPrep, Weaver and Company, Aurora, CO) and impedance was measured. If impedance levels were higher than 20kOhms, the skin was scrubbed again and signals were checked to ensure that EMG signal peaks were discernible from noise before beginning the task.

Data was recorded using a BIOPAC system with an MP150 amplifier (gain: 2000) and AcqKnowledge software (BIOPAC Systems Inc., Goleta, CA). The unfiltered EMG channel was acquired and sampled at a rate of 1000 Hz. Following data recording, files were manually inspected to check for excessive noise and clarity of signal (57 files or 8.4% removed due to bad signal/equipment recording problems). The remaining files were then processed by applying a digital FIR band pass filter (28Hz – 500Hz) and the average rectified value was obtained using AcqKnowledge's "Derive Average Rectified EMG" procedure with a 25ms moving window. Startle probes were automatically identified from the digital stimulus channel input, and startle responses for each probe were derived by

subtracting the average EMG magnitude in a 50ms window before the probe onset from the maximum value in the +20ms to +150ms post-probe window.

2.2.3 Data cleaning.—Individual trials (i.e., startle probe administrations) were considered to be contaminated by baseline noise and removed from the scored file if the standard deviation of the pre-probe baseline was greater than three times the average standard deviation of all pre-probe baselines for that individual. Non-response trials were also removed if the maximum EMG magnitude during the post-probe window was less than one standard deviation above the pre-probe baseline magnitude for that trial. Participants were excluded from further analysis if they quit the task before the acquisition phase, if more than 20% of their SRs from the task were excluded for noisy baselines or non-response, or if their average SR magnitude across the task was greater or less than three times the standard deviation of the mean SR across participants within the same site, resulting in a total of 49 participants (7.3%) being removed from the analyses.

2.3 Data Analysis

2.3.1 Startle response metrics.—To compare the impact of different transformations of the SR, we created several variations of startle response metrics. These used both the raw EMG magnitudes for each startle probe response and the standardized magnitudes in which each SR was normalized to a within-person T-score distribution with a mean of 50 and standard deviation of 10. For both raw and standardized values, a score for each of the CS+, CS-, and ITI stimuli in each phase was derived as an average across the 4–10 startle probes, leaving out the first probe in the CS+ stimulus to allow opportunity for acquisition to occur. The metrics used for analysis are as follows:

1. The overall raw SR, averaged across all CS+, CS-, and ITI probes in the task (“Overall Raw”)
2. The means of the CS+ or CS- startle probes, calculated separately for raw (“Raw No Differential”) and standardized values (“T-score No Differential”)
3. A differential measure of the *potentiated (FPS)* and *unconditioned* startle in which the mean of the ITI startle probes was subtracted from the means of the CS+ and CS- probes, respectively - also calculated for raw (“Raw Differential”) and standardized values (“T-score Differential”)

2.3.2 Reliability analysis.—For each SR metric described above, we estimated test-retest reliability by calculating Pearson correlations between scores assessed at visit 1 and at visit 2.

2.3.3 Heritability analysis.

2.3.3.1 Univariate twin models. Standard biometrical twin modeling (Neale & Cardon, 1992) decomposes the observed variance in SR metrics into additive genetic (*A*), common environment (*C*), and unique environment (*E*) latent factors. *A* reflects the additive effects of all genetic loci and contributes twice as much to the MZ versus the DZ correlation, since MZ twins share 100% of their segregating genetic variants, whereas DZ twins share, on average, 50%. *C* reflects aspects of the environment that make twins raised in the same

family more alike than random pairs of individuals and contributes equally to the MZ and DZ correlations. E reflects aspects of the environment that are unique to an individual (plus measurement error) and is uncorrelated between twins in a pair. In the typical model fitting sequence, the within-pair correlation for each phenotype is decomposed into A , C , and E factors (ACE /full model), and then individual parameters are constrained to zero (i.e., CE model, AE model, E model) to test for significant contributions of each of these parameters to the trait variance. Sub-models are compared to the full model via difference in $-2 \times \log$ -likelihood ($-2LL$) statistics, which follows a χ^2 distribution. Therefore, a significant p -value for the associated χ^2 statistic indicates a significant deterioration in fit to the data and suggests that the sub-model should be rejected.

2.3.3.2 Multivariate twin model.: A multivariate (Cholesky) decomposition extends the univariate twin analysis by incorporating all relevant phenotypes into a single biometrical model. The logic behind the multivariate model is much the same as that of the univariate models but additionally includes cross-trait covariance paths which are based on cross-twin cross-trait correlations and allow for the quantification of genetic and/or environmental influences that are common to two or more phenotypes. This approach was used to distinguish genetic and environmental factors that are unique to modulated startle while controlling for influences that are shared with baseline SR.

All biometric analyses were performed in the R statistical environment (R Core Team, 2015) using the OpenMx package (Neale et al., 2015). Site of data collection (VCU or NIMH), age, and sex were used as covariates in all models. These covariates were associated with mean differences in raw EMG values; for example, for overall SR, older age ($\beta=0.10$, 95% confidence interval [CI]: 0.02–0.19), female sex ($\beta=0.53$, 95% CI: 0.31–0.76), and NIMH site ($\beta=0.67$, 95% CI: 0.33–1.02) were linked to higher raw EMG values, although these effects were not significant predictors of the standardized response scores.

2.3.4 Simulation analysis.—Given the limited size of our twin sample, we expected that we might be underpowered to reliably estimate parameters in the biometric models. Further, the specific attributes of this sample and this paradigm might limit wider generalizability of results. We therefore conducted a series of data simulations to compare the effects of using the aforementioned metrics of the SR on heritability estimates in a simulated sample. This allowed us to manipulate the parameters to model different plausible conditions (e.g. higher or lower heritability; higher or lower dependence of FPS on baseline startle magnitude) in a larger and more statistically powerful scenario and thereby obtain more reliable and generalizable results. We generated multiple datasets with 1000 pairs each of MZ and DZ twins using the MASS package (Venables & Ripley, 2002) in R. In each dataset we varied the range of heritability of baseline and potentiated startle (0.3 to 0.8) as well as their phenotypic correlation (0.3 to 0.9). For simplicity of interpretation, we included only two variables: a baseline and CS+ potentiated measure. Baseline startle and FPS were simulated under a correlated factor model, whereby two latent factors were presumed to underlie two sets (baseline and FPS, respectively) of five startle probes each. Varying degrees of measurement error in each probe were also simulated, and the correlation

between baseline and FPS probes was driven by a correlation between their underlying latent factors (see Figure 1).

Individual probes were simulated to have factor loadings with a random value sampled from a uniform distribution between 0.80 and 0.95 (with the remaining 0.05–0.20 variance due to individual measurement error). FPS probes had a mean value either 130% or 300% of the baseline. From this raw simulated data, we applied the same series of measurement comparisons (raw/T-score no differential, raw/T-score differential scores, and a bivariate model including baseline and FPS). We then compared the parameter estimates returned from the univariate or multivariate biometric models of these variables to the values that were simulated in order to identify whether transforming the measures resulted in biased heritability estimates. Simulations were repeated 1000 times for each set of parameters, and the estimates returned from the models were averaged across these 1000 simulations.

3. Results

3.1 Descriptive Statistics and Reliability

Table 1 displays descriptive statistics, twin correlations, and visit 1-visit 2 correlations for all SR metrics used in the current analyses. For both raw and standardized scores, higher values were found for the CS+ probes relative to the CS- probes, indicative of successful conditioned fear learning. For the raw scores, the correlation was higher for MZ twins relative to DZ twins for the overall SR and all No Differential measures, indicative of genetic influences. This was also generally true for the T-score standardized No Differential measures, although the overall magnitudes of the correlations were attenuated and mostly not significant. Twin correlations were further attenuated in the Differential scores in which ITI startle magnitude was subtracted from the CS+/CS- response magnitude, with overall correlations near zero and larger correlations for DZ relative to MZ twins for three of the eight metrics. Similarly, the test-retest reliability was moderate ($r = .54-.58$, p 's $< 2 \times 10^{-8}$) for all raw unstandardized measures but poor for all standardized and differential measures ($r < .25$).

3.2 Univariate Heritability

Table 2 presents the parameter estimates for the univariate ACE models for each startle metric. Full model fitting results can be found in Supplementary Tables 1–17. For each metric, the best-fitting model was selected by constraining the A and/or C parameters to zero (i.e., testing overall effects of familial aggregation) and identifying the model that did not exhibit a significant decrease in fit of the loglikelihood; if A and C could be dropped individually but not jointly, the full ACE model was retained. Based on this, Overall Raw mean score was moderately heritable ($A = 0.34$) with the remaining variance due to shared and unique environment ($C = 0.12$; $E = 0.54$). Although the confidence intervals for A and C included zero, dropping both of these parameters led to a highly significant decrease in model fit ($p = 1 \times 10^{-8}$), so evidence for familial aggregation was supported. The raw No Differential CS+ scores were modestly to moderately heritable, with A estimates of 0.06 to 0.35 for the acquisition and extinction phases, respectively. The raw No Differential CS- scores were modestly heritable in the extinction phase ($A = 0.21$) but not the acquisition

phase ($A = 0.00$). All standardized No Differential scores and both the raw and standardized Differential scores were accounted for entirely by unique environment, as the A and C parameters could be dropped from these models without a decrease in fit ($p > .05$).

3.3 Multivariate heritability

The results from the multivariate Cholesky biometric model are shown in Table 3. In the top panel, phenotypic (i.e., within-person) correlations were very high ($r > 0.78$) between SR measures throughout the task. Path estimates, shown in the middle panel, represent the contributions of latent genetic/environmental factors to each measure, with cross-paths on the off-diagonals indicating whether those latent influences are shared between measures. A simplified diagram of the model is presented in Figure 2. The first latent factor, in the left-most column, loads onto all measures and reflects influences common across all of them. The second factor shows influences common to all measures except the first (i.e., except Habituation/baseline startle), and so on. For additive genetics (A), the first factor had significant loadings on SR across all phases and stimuli, indicative of a single common heritable factor underlying all SR measures. There were modest path loadings on the second genetic factor (0.07–0.19) - encompassing influences unique to the fear-conditioned modulation of the SR - but these estimates did not significantly differ from zero. Similarly, for the common environment (C), a single latent factor influenced all SR outcomes, with no evidence for influences unique to modulated startle. Unique environmental (E) factors had the greatest contribution to all SR outcomes, with multiple sets of influences that were shared across all SR outcomes as well as unique to modulated responses and to each specific phase/stimulus in the task. Unique E influences are expected since they reflect uncorrelated sources of error between the measures.

The lower panel of Table 3 presents the proportions of variance (diagonal elements) and covariance between measures (off-diagonals) attributable to A , C , and E , which represent the aggregate effects of all the latent factor contributions from each of these sources to the variance of the measure or the covariance between two measures. These results demonstrate that 18–32% of the inter-individual variability in SR in this sample is attributable to additive genetics, 8–19% to the common environment, and 54–59% to the unique environment, with similar proportions for the covariance. Together with the path estimates from this model, it can be inferred that A and C have significant influences on SR, but this is almost entirely through factors that are common across all of the highly-correlated measures with little to no specific influence on modulation of the SR.

3.4 Model Simulation

Results from the simulation analyses are shown in Table 4. The model estimates in the three right-hand columns demonstrate the estimated heritability of FPS when it is defined as a (1) raw or (2) T-score standardized difference score calculated by subtracting out the neutral/ITI response, or (3) when the covariance between baseline and FPS is explicitly specified in a bivariate twin model. Comparison of these estimates with the simulated parameter of the heritability that is unique to FPS (i.e., excluding what is shared with baseline; bolded center column in Table 4) is indicative of any biases in the heritability estimates caused by statistical transformations of the different metrics. Both raw and standardized difference

scores returned substantially biased heritability results from the simulated values under multiple conditions. Specifically, when the phenotypic correlation between baseline and FPS was very high (e.g. 0.90, as observed in our empirical study), univariate heritability models of the differences scores typically overestimated the unique FPS heritability by 20–30%. Conversely, when the phenotypic correlation was low but the heritability of FPS was higher than that of baseline, FPS-unique heritability was sometimes underestimated. The direction and degree of bias was highly dependent on the architecture of the two traits. Estimates from the raw versus T-score standardized differential were similar when the potentiation effect was small (130% of baseline) but diverged when the magnitude of fear potentiation was large (300%). The bivariate model jointly estimating the effects of raw baseline and potentiated scores consistently returned unbiased estimates of the FPS-unique heritability across all conditions.

4. Discussion

4.1 Summary of Findings

In the current study, we sought to evaluate the utility of FPS as an endophenotype by establishing the requisite criterion of its heritability and reliability in pre-adolescents. We also investigated the impact of measurement transformations relevant to FPS and other conditional phenotypes on the estimates of heritability in both real and simulated data. Overall, the univariate and multivariate twin models indicated that measures of raw, untransformed SR throughout a fear conditioning task were moderately stable across a one-month period and modestly heritable, but these heritable influences were driven by a common genetic effect shared with baseline startle response. Virtually no evidence of heritability was apparent when these basal influences were controlled for by the use of a differential score, within-person standardization, or joint multivariate model. The poor test-retest reliability of the FPS metrics also indicates that such measures effectively capture only noise; this high measurement error limits the upper bound of heritable influences on a trait that can be identified even in ideal analytic scenarios. Simulations indicated that, under situations of true FPS-specific genetic influences, a multivariate model would provide less biased estimates compared with univariate models of transformed measures. It can be inferred that studies of other types of modulated startle, or of conditional phenotypes with similar attributes, will return biased heritability estimates if univariate twin models are applied to (standardized) difference scores rather than a multivariate model.

4.2 Comparison with Previous Studies

The current findings are in line with those of two previous twin studies (Anokhin et al., 2007; Dhamija et al., 2017) which concluded that a common pathway model with a single set of genetic and environmental factors was the driving force behind startle response to neutral, positive, and negative affective stimuli. Another (Vaidyanathan et al., 2014) found that overall raw SR, but not modulated startle, was moderately heritable. Our sample of 9- to 14-year-olds did not implicate substantial differences in the genetic architecture of overall SR during this developmental stage compared to results from the previous studies of adolescents and adults. Similarly, we can conclude that the familial influences on fear

potentiation of the SR are small or at least indistinguishable from their modest but significant influences on basal startle.

Though relatively consistent with other twin studies of SR, the results of the current study diverged from a previous analysis that used a conditioning paradigm and multivariate twin model and found unique heritability of the fear-conditioned response (Hettema et al., 2003). That study differed in its use of skin conductance response rather than SR, its adult sample, and the use of directly fear-relevant stimuli. However, it does indicate that genetic influences on fear conditioning processes exist, and the question of whether or not these same influences affect FPS remains an open one.

4.3 Implications for Genetic Research

One possible explanation for our inability to detect FPS-specific influences is that a stronger potentiator might be required to differentiate FPS from baseline SR. We found very high correlations between startle responses in all phases/stimuli of the task, and the simulations indicated that the unique heritability of potentiated (modulated) startle is only ~5–15% in conditions where the correlation with basal startle is high, regardless of total heritability. This means that there is very little remaining phenotypic variance from which to parse genetic versus environmental influences after controlling for baseline effects, no matter which method is used for this control. High phenotypic correlations may be a result of large shared genetic influences on the electrical activity of muscles as well as environmental factors such as the technical conditions that are shared between all startle probes for a single person during the same experimental setup. These factors may overwhelm any smaller contributions to the modulation of startle response, necessitating very large samples to detect them. Post-hoc power analyses indicated that we had moderate power (65–77%) to detect genetic influences accounting for ~40% of the variance in a univariate model with our current sample size; however, this was substantially lower (0–33%) when accounting for the high phenotypic correlation with basal startle and correspondingly lower unique heritability in a multivariate model, with power depending on the true heritability of both outcomes. If potentiated startle can be more reliably separated from baseline, perhaps it will be possible to distinguish those factors that specifically affect its modulation. The higher test-retest reliability (albeit for a shorter interval) seen for modulation of startle in a more aversive unpredictable threat-of-shock task (Kaye, Bradford, & Curtin, 2016) speaks to the possibility that a stronger potentiator may improve the utility of FPS measures. Unfortunately, such paradigms are more unpleasant and therefore more difficult to apply in research settings, especially with children or other vulnerable populations. Other ways to maximize effect size and biological relevance, such as using aggregate measures of multiple physiological biomarkers/endophenotypes, should be considered.

Beyond the implications for the endophenotypic characterization of FPS that can be drawn from the specific FPS paradigm we employed, this study uncovers some important findings about the broader use of differential scores in twin modeling studies. We expected that the loss of inter-individual variability from standardizing scores would result in an underestimate of the heritability of FPS. Although this was not exactly true, such a method still showed bias (upwards and downwards) in the estimates of heritability. We suspect that

the use of difference scores caused some of the statistical “heritability” of one co-dependent phenotype to shift into the other due to their correlated values, in different directions depending on the heritability of each and strength of the phenotypic correlation. Creating difference scores means that the error associated with both terms is combined, contributing to greater – and unpredictable – noise (nuisance variance) in the outcome scores. The results highlight the importance of applying appropriate statistical models and of explicitly including the covariance between co-dependent traits when differential scores (e.g. stimulus-response; treatment effects; change in a trait across time) are the phenotype of interest in a genetic study design. Interestingly, however, the simulations implied that a high correlation between measures would lead to overestimates of the heritability in differential scores, but the estimates of essentially zero heritability found here do not reflect that prediction. In this study, and in others (Anokhin et al. 2007; Dhamija et al., 2017; Vaidyanathan et al., 2014), the twin correlations were often negative and frequently larger in DZ than MZ pairs. We speculate that small potentiation/modulation effects of the paradigms used and high correlations with baseline startle renders the differential scores effectively “noise”, capturing little meaningful signal that can be observed in the twin correlations.

4.4 Limitations

While this study addresses several key gaps within the startle heritability literature, it also suffers from some potential limitations. Several of these limitations stem from using a pre-adolescent sample. First, while the small startle potentiation observed is not directly related to age differences in biology, it may be related to the lower level of discomfort presented in the current paradigm designed for this younger sample (as opposed to adult paradigms using more uncomfortable startle probes such as shocks). Second, when assessing children in this age range, it can be difficult to obtain good EMG recordings due to their inability to sit still for long periods and maintain attention to task. Also, participants with the strongest fear reaction may also have been the ones to quit or refuse the task altogether, further reducing the ability to discern heritability of FPS response from baseline response. This might cause differences in the patterns observed in our study compared to samples with clinically significant levels of fear/fearlessness. This sample is also relatively small for a twin study, although we inform our findings with the use of data from a large simulated sample. Other measurement issues such as habituation of response and learning effects (i.e., whether or not participants successfully identified the CS+ versus CS-) might play a role, although we found similar twin model results in post-hoc analyses when using only the last half of probes in the task (post-habituation) as well as when removing individuals who reported they could not identify which stimulus predicted the UCS (non-learners). An additional limitation for FPS is that “baseline” measures may actually be capturing some level of fear potentiation as participants may be in a high-arousal anticipatory state throughout the entirety of the task, having been informed that it will be aversive. If so, there could be artificial inflation in the correlation between baseline and FPS. On the other hand, participants were in a controlled laboratory setting with no real threat of harm or danger and were able to stop the task at any time and therefore may not have acquired truly fearful responses. In contrast, animal models in which the animals have no such cognitive safety cues have shown much higher heritability of FPS (McCaughan et al., 2000).

4.5 Conclusions

Despite the limitations mentioned, there are several important takeaways from this study. This was the first study to examine the heritability of FPS in pre-adolescents. We reach a similar conclusion as other studies of affect-modulated startle in older samples: that baseline/overall startle is heritable but the heritability of FPS remains less clear. We propose a few recommendations for the future use of modulated startle response measures. First, any studies using these measures should utilize a paradigm in which FPS can be measured reliably (else costly genetic investigations are unwarranted), dependably separated from baseline SR (for example, using a more aversive/unpredictable threat stimulus to potentiate startle and a baseline measure that is as neutral as possible) and which captures true differences in threat/emotion processing systems engaged between a resting condition and a fear-negative affect-inducing condition. While this is certainly difficult to achieve given ethical and practical constraints, it is necessary in order to determine whether the lack of observed heritability is due to true (total) overlap with the genetic influences on basal startle or merely a result of confounding factors between them. Virtual reality technology may be one possible means to create an ecologically valid environment for this assessment. Second, although controlling for inter-individual differences in baseline startle is important for making comparisons of FPS, the use of a multivariate model in which both facets are included is preferable to a statistical transformation to remove differences. This recommendation is broadly applicable for studies, especially genetic epidemiology studies, investigating conditional phenotypes that are inherently defined by their relationship to a baseline level.

The evidence thus far does not support the utility of FPS as an endophenotype for understanding the genetic and biological influences on threat-response systems and related psychopathology. Though now indicated by multiple studies, this conclusion is surprising given the robust literature demonstrating associations between FPS and fear-relevant traits as well as the position of FPS in a cohesive theoretical network of biological threat-response systems. The poor psychometric properties exhibited by FPS in this study indicate cause for concern in the previous literature (and a potential explanation for inconsistencies between studies), although it is possible that the application of stronger modulators or the restriction to clinical samples in some previous studies resulted in more reliable startle metrics. The numerous replicated direct links of FPS to clinical conditions and dispositional traits (Lake et al., 2011; Lang, 1995; Lissek, Powers et al., 2005; Vaidyanathan et al., 2009), its evolutionary conservation (Grillon & Bass, 2003), and its heritability in other species (McCaughan et al., 2000) suggest that some critical piece may yet be missing from the puzzle.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The VCU-JAS study was supported by the National Institutes of Health (R01MH098055 to JMH and NIMH-IRP-ziamh002781 to DSP). The Mid-Atlantic Twin Registry is supported through the NIH Center for Advancing

Translational Research Grant Number UL1TR000058. JES, AAM, CSR, and JLB were supported by the NIH (T32MH020030).

References

- Anokhin AP, Golosheykin S, & Heath AC (2007). Genetic and environmental influences on emotion-modulated startle reflex: A twin study. *Psychophysiology*, 44, 106–112. 10.1111/j.1469-8986.2006.00486.x [PubMed: 17241146]
- 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. 10.1111/j.1469-8986.2005.00271.x [PubMed: 15720576]
- Bradford DE, Starr MJ, Shackman AJ, & Curtin JJ (2015). Empirically based comparisons of the reliability and validity of common quantification approaches for eyeblink startle potentiation in humans. *Psychophysiology*, 52(12), 1669–1681. 10.1111/psyp.12545 [PubMed: 26372120]
- Britton JC, Grillon C, Lissek S, Norcross MA, Szuhany KL, Chen G, . . . Pine DS (2013). Response to learned threat: An fMRI study in adolescent and adult anxiety. *American Journal of Psychiatry*, 170, 1195–1204. 10.1176/appi.ajp.2013.12050651 [PubMed: 23929092]
- Carney DM, Moroney E, Machlin L, Hahn S, Savage JE, Lee M, . . . Hettema JM (2016). The twin study of negative valence emotional constructs. *Twin Research and Human Genetics*, 19(5): 456–464. 10.1017/thg.2016.59 [PubMed: 27457271]
- Cuthbert BN (2014). Translating intermediate phenotypes to psychopathology: The NIMH Research Domain Criteria. *Psychophysiology*, 51(12), 1205–1206. 10.1111/psyp.12342 [PubMed: 25387702]
- Dhamija D, Tuvblad C, Dawson ME, Raine A, & Baker LA (2017). Heritability of startle reactivity and affect modified startle, *International Journal of Psychophysiology*, 115, 57–64. 10.1016/j.ijpsycho.2016.09.009. [PubMed: 27666795]
- Edwards JR (2011). Ten difference score myths. *Organizational Research Methods*, 4(3), 265–287. 10.1177/109442810143005
- Flint J, & Munafo MR (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine*, 37(2), 163–180. 10.1017/S0033291706008750 [PubMed: 16978446]
- Glenn CR, Klein DN, Lissek S, Britton JC, Pine DS, & Hajcak G (2012). The development of fear learning and generalization in 8–13 year-olds. *Developmental Psychobiology*, 54(7), 675–684. 10.1002/dev.20616 [PubMed: 22072276]
- Glenn CR, Lieberman L, & Hajcak G (2012). Comparing electric shock and a fearful screaming face as unconditioned stimuli for fear learning. *International Journal of Psychophysiology*, 86(3), 214–219. 10.1016/j.ijpsycho.2012.09.006 [PubMed: 23007035]
- Grillon C, & Bass J (2003). A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Child Neurophysiology*, 114, 1557–1579. 10.1016/S1388-2457(03)00202-5
- Gottesman II, & Gould TD (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*, 160, 636–645. 10.1176/appi.ajp.160.4.636 [PubMed: 12668349]
- Hamm AO, & Weike AI (2005). The neuropsychology of fear learning and fear regulation. *International Journal of Psychophysiology*, 57(1), 5–14. 10.1016/j.ijpsycho.2005.01.006 [PubMed: 15935258]
- Hettema JM, Annas P, Neale MC, Kendler KS, & Fredrikson M (2003). A twin study of the genetics of fear conditioning. *Archives of General Psychiatry*, 60(7), 702–708. 10.1001/archpsyc.60.7.702 [PubMed: 12860774]
- Jackson ED, Payne JD, Nadel L, & Jacobs WJ (2006). Stress differentially modulates fear conditioning in healthy men and women. *Biological Psychiatry*, 59(6), 516–522. 10.1016/j.biopsych.2005.08.002 [PubMed: 16213468]
- Kaye JT, Bradford DE, & Curtin JJ (2016). Psychometric properties of startle and corrugator response in NPU, Affective Picture Viewing, and Resting State tasks. *Psychophysiology*, 53(8), 1241–1255. 10.1111/psyp.12663 [PubMed: 27167717]

- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, & Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602. 10.1001/archpsyc.62.6.593 [PubMed: 15939837]
- Lau JYF, Lissek S, Nelson EE, Lee Y, Roberson-Nay R, Poeth K, Jennes J Pine DS (2008). Fear conditioning in adolescents with anxiety disorders: Results from a novel experimental paradigm. *Journal of American Academy of Child and Adolescent Psychiatry*, 47, 94–102. 10.1097/chi.0b01e31815a5f01
- Lake AJ, Baskin-Sommers AR, Li W, Curtin JJ, & Newman JP (2011). Evidence for unique threat-processing mechanisms in psychopathic and anxious individuals. *Cognitive, Affective, and Behavioral Neuroscience*, 11, 451–462. 10.3758/s13415-011-0041-2
- Lang PJ (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5), 372–385. <http://psycnet.apa.org/doi/10.1037/0003-066X.50.5.372> [PubMed: 7762889]
- Lilley EC, & Silberg JL (2013). The Mid-Atlantic Twin Registry, revisited. *Twin Research and Human Genetics*, 16, 424–428. 10.1017/thg.2012.125 [PubMed: 23218199]
- Lissek S, Baas JMP, Pine DS, Orme K, Dvir S, Nugent M, . . . Grillon C (2005). Airpuff startle probes: An efficacious and less aversive alternative to white-noise. *Biological Psychology*, 68(3), 283–297. 10.1016/j.biopsycho.2004.07.007 [PubMed: 15620795]
- Lissek S, Orme K, McDowell DJ, Johnson LL, Luckenbaugh DA, Baas JM, . . . Grillon C (2007). Emotion regulation and potentiated startle across affective picture and threat-of-shock paradigms. *Biological Psychology*, 76(1), 124–133. 10.1016/j.biopsycho.2007.07.002 [PubMed: 17692453]
- Lissek S, Powers AS, McClure EB, Phelps EA, Woldehawariat G, Grillon C, & Pine DS (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy*, 43, 1391–1424. 10.1016/j.brat.2004.10.007 [PubMed: 15885654]
- Loomans MM, Tulen JH, & van Marle HJ (2015). The startle paradigm in a forensic psychiatric setting: Elucidating psychopathy. *Criminal Behaviour and Mental Health*, 25, 42–53. 10.1002/cbm.1906 [PubMed: 24664972]
- Lynam DR, Charnigo R, Moffitt TE, Raine A, Loeber R, & Stouthamer-Loeber M (2009). The stability of psychopathy across adolescence. *Development and Psychopathology*, 21(4), 1133–1153. 10.1017/s0954579409990083 [PubMed: 19825261]
- McCaughran JA Jr., Bell III J, & Hitzemann RJ (2000). Fear-potentiated startle response in mice: Genetic analysis of the C57BL/6J and DBA/2J intercross. *Pharmacology Biochemistry and Behavior*, 65, 301–312. 10.1016/S0091-3057(99)00216-6
- Neale M, & Cardon L (1992). *Methodology for genetic studies of twins and families*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Neale MC, Hunter MD, Pritikin JN, Zahery M, Brick TR, Kirkpatrick RM, . . . Boker SM (2015). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 81, 535–549. 10.1007/s11336-014-9435-8 [PubMed: 25622929]
- Patrick CJ (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology*, 31(4), 319–330. 10.1111/j.1469-8986.1994.tb02440.x [PubMed: 10690912]
- Quevedo K, Smith T, Donzella B, Schunk E, & Gunnar M (2010). The startle response: developmental effects and a paradigm for children and adults. *Developmental Psychobiology*, 52(1), 78–89. 10.1002/dev.20415 [PubMed: 20027622]
- R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Savage JE, Sawyers C, Roberson-Nay R, & Hettema JM (2017). The genetics of anxiety-related negative valence system traits. *American Journal of Medical Genetics Part B*, 174(2), 156–177. 10.1002/ajmg.b.32459
- Vaidyanathan U, Malone SM, Miller MB, McGue M & Iacono WG (2014). Heritability and molecular genetic basis of acoustic startle eye blink and affectively modulated startle response: A genome-wide association study. *Psychophysiology*, 51, 1285–1299. 10.1111/psyp.12348 [PubMed: 25387708]
- Vaidyanathan U, Patrick CJ, & Cuthbert BN (2009). Linking dimensional models of internalizing psychopathology to neurobiological systems: Affect-modulated startle as an indicator of fear and

distress disorders and affiliated traits. *Psychological Bulletin*, 135(6), 909–942. 10.1037/a0017222 [PubMed: 19883142]

Venables WN & Ripley BD (2002). *Modern Applied Statistics with S. Fourth Edition* New York, NY: Springer ISBN 0–387-95457–0

Yeomans JS, Li L, Scott BW, & Frankland PW (2002). Tactile, acoustic and vestibular systems sum to elicit the startle reflex. *Neuroscience and Biobehavioral Reviews*, 26(1), 1–11. 10.1016/S0149-7634(01)00057-4 [PubMed: 11835980]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Schematic of the correlated latent factor model used to simulate the relationship between baseline and potentiated startle while varying the factor loadings (f) of each simulated probe response, their corresponding error (e) terms, the difference in means (m), and the correlation between baseline and potentiated startle (r).

Figure 2.

Partial schematic of a Cholesky model showing the decomposition of covariance between multiple measures of the startle response task as a function of genetic influences (A) shared between variables. The full model includes identical sets of pathways for common environmental (C) and unique environmental (E) influences, which are estimated by comparing phenotypic correlations between twins. Acq = acquisition phase, Ext = extinction phase.

Table 1.

Descriptive Statistics and Twin and Visit 1-Visit 2 (v1-v2) Correlations for Startle Metrics

Startle Metric	Raw Score				Standardized Score							
	M	SD	r _{Total}	r _{MZ}	r _{DZ}	r _{v1-v2}	M	SD	r _{Total}	r _{MZ}	r _{DZ}	r _{v1-v2}
Overall	2.24	1.37	.38***	.53***	.28**	.58***	—	—	—	—	—	−.08
No Differential												
Acquisition phase - CS+	2.34	1.47	.36***	.52***	.23**	.52***	50.70	4.17	.08	.18	.02	−.07
Acquisition phase - CS−	2.19	1.39	.39***	.42***	.36***	.57***	48.83	3.95	.03	.09	−.03	.07
Extinction phase - CS+	1.98	1.32	.32***	.40**	.26**	.55***	47.69	3.61	−.02	.07	−.05	.02
Extinction phase - CS−	1.94	1.30	.37***	.48***	.31***	.53***	46.81	3.71	.07	.16*	.02	.03
Differential												
Acquisition phase - CS+	0.29	0.59	.17*	.10	.21*	.11	2.90	5.20	.06	.10	.04	.07
Acquisition phase - CS−	0.14	0.51	−.03	−.01	−.05	.16	1.27	4.77	−.03	.05	−.09	.14
Extinction phase - CS+	0.24	0.51	−.02	−.12	.02	.24**	2.40	4.36	−.02	−.12	.01	.07
Extinction phase - CS−	0.17	0.49	.15*	.20*	.13	.02	1.56	4.30	.07	.10	.06	.07

Note: Overall = mean raw EMG magnitude of responses to all startle probes throughout the task; No Differential = mean raw EMG magnitude of startle responses during each stimulus; Differential = mean raw EMG magnitude of startle responses during each stimulus, subtracting out responses during a neutral ITI period; MZ = monozygotic twins; DZ = dizygotic twins.

* p < 0.05

**

p < 0.01

p < 0.0001

Univariate biometric model estimates of additive genetic (*A*), common environmental (*C*), and unique environmental (*E*) contributions to startle response as characterized by different metrics.

Table 2.

Startle Metric	Raw Score			Standardized Score		
	Variance Proportion(95% CIs)			Variance Proportion(95% CIs)		
	<i>A</i>	<i>C</i>	<i>E</i>	<i>A</i>	<i>C</i>	<i>E</i>
Overall	0.34(0.00–0.63)	0.12(0.00–0.43)	0.54(0.42–0.72)	–	–	–
<i>No Differential</i>						
Acquisition phase - CS+	0.35(0.00–0.59)	0.07(0.00–0.41)	0.59(0.45–0.78)	0.15(0.00–0.36)	0.00(0.00–0.23)	0.85(0.65–1.00)
Acquisition phase - CS–	0.00(0.00–0.37)	0.35(0.06–0.49)	0.66(0.51–0.78)	0.04(0.00–0.23)	0.00(0.00–0.16)	0.96(0.77–1.00)
Extinction phase - CS+	0.06(0.00–0.51)	0.27(0.00–0.47)	0.67(0.50–0.84)	0.01(0.00–0.24)	0.00(0.00–0.15)	0.99(0.76–1.00)
Extinction phase - CS–	0.21(0.09–0.78)	0.22(0.00–0.49)	0.58(0.43–0.77)	0.14(0.00–0.38)	0.00(0.00–0.24)	0.86(0.64–1.00)
<i>Differential</i>						
Acquisition phase - CS+	0.00(0.00–0.24)	0.09(0.00–0.22)	0.91(0.75–1.00)	0.14(0.00–0.28)	0.00(0.00–0.17)	0.86(0.72–1.00)
Acquisition phase - CS–	0.00(0.00–0.10)	0.00(0.00–0.07)	1.00(0.87–1.00)	0.00(0.00–0.15)	0.00(0.00–0.09)	1.00(0.83–1.00)
Extinction phase - CS+	0.00(0.00–0.18)	0.00(0.00–0.12)	1.00(0.81–1.00)	0.00(0.00–0.19)	0.00(0.00–0.13)	1.00(0.79–1.00)
Extinction phase - CS–	0.20(0.07–0.32)	0.00(0.00–0.28)	0.80(0.62–1.00)	0.12(0.00–0.35)	0.00(0.00–0.22)	0.88(0.66–1.00)

Note: Overall = mean raw EMG magnitude of responses to all startle probes throughout the task; No Differential = mean raw EMG magnitude of startle responses during each stimulus; Differential = mean raw EMG magnitude of startle responses during each stimulus, subtracting out responses during a neutral ITI period; CI = confidence interval; Standardized = transformed via within-person T-score standardization.

Table 3.

Multivariate associations between multiple measures of startle response in a fear conditioning paradigm.

		Habituation	Acq CS+	Acq CS–	Ext CS+	Ext CS–
Phenotypic Correlations	Habituation	--				
	Acq CS+	0.87***	--			
	Acq CS–	0.88***	0.93***	--		
	Ext CS+	0.80***	0.89***	0.88***	--	
	Ext CS–	0.79***	0.89***	0.89***	0.93***	--
Standardized Path Estimates	<i>A</i>	Habituation	0.56*			
		Acq CS+	0.56*	0.10		
		Acq CS–	0.46*	0.09	0.00	
		Ext CS+	0.42*	0.07	0.00	0.00
		Ext CS–	0.44*	0.19	0.00	0.00
	<i>C</i>	Habituation	0.38*			
		Acq CS+	0.29*	0.00		
		Acq CS–	0.44*	0.00	0.00	
		Ext CS+	0.42*	0.00	0.00	0.00
		Ext CS–	0.43*	0.00	0.00	0.00
Proportions of (Co)variance	<i>E</i>	Habituation	0.73*			
		Acq CS+	0.65*	0.42*		
		Acq CS–	0.67*	0.23*	0.31*	
		Ext CS+	0.60*	0.31*	0.03	0.43*
		Ext CS–	0.59*	0.28*	0.05	0.26*
	<i>A</i>	Habituation	0.32*			
		Acq CS+	0.35*	0.32*		
		Acq CS–	0.28*	0.29*	0.22*	
		Ext CS+	0.28*	0.27*	0.23*	0.18*
		Ext CS–	0.29*	0.30*	0.25*	0.21*
<i>C</i>	<i>C</i>	Habituation	0.15*			
		Acq CS+	0.12*	0.08		
		Acq CS–	0.18*	0.14*	0.19*	

		Habituation	Acq CS+	Acq CS-	Ext CS+	Ext CS-
	Ext CS+	0.19 *	0.14 *	0.21 *	0.17 *	
	Ext CS-	0.20 *	0.14 *	0.21 *	0.19 *	0.18 *
<i>E</i>	Habituation	0.54 *				
	Acq CS+	0.53 *	0.59 *			
	Acq CS-	0.53 *	0.57 *	0.59 *		
	Ext CS+	0.53 *	0.59 *	0.56 *	0.65 *	
	Ext CS-	0.51 *	0.56 *	0.54 *	0.60 *	0.59 *

Note: Within-person phenotypic correlations are shown in the top panel. Standardized path estimates of additive genetic (*A*), common environmental (*C*), and unique environmental (*E*) contributions are in the top panel, while proportions of (co)variance from each of these sources are in the lower panel. All ACE estimates are derived from the multivariate Cholesky model (Figure 2). *Acq* = acquisition phase, *Ext* = extinction phase.

*
p<0.05

p<0.0001

Table 4.

Heritability results estimated from different metrics of startle response in simulated data (1000 datasets each with 1000 pairs of monozygotic and 1000 pairs of dizygotic twins) and comparison with the simulated parameters for baseline (BSL) and fear-potentiated startle (FPS) under different models of shared and unique heritability.

FPS = 130% of baseline						
Simulated parameters			Estimates returned for FPS heritability			
BSL-FPS phenotypic correlation	BSL total heritability	FPS total heritability	FPS unique heritability	Raw difference score	Standardized difference score	Bivariate model of raw BSL and FPS
0.90	0.80	0.70	0.10	[*] 0.43	[*] 0.37	0.11
0.60	0.80	0.70	0.48	[*] 0.74	0.51	0.46
0.30	0.80	0.70	0.67	[*] 0.81	[*] 0.79	0.64
0.90	0.80	0.55	0.06	[*] 0.37	[*] 0.35	0.07
0.60	0.80	0.40	0.30	[*] 0.71	[*] 0.70	0.29
0.30	0.80	0.40	0.40	[*] 0.75	[*] 0.74	0.38
0.90	0.50	0.70	0.12	[*] 0.45	[*] 0.39	0.13
0.60	0.50	0.70	0.15	0.19	0.19	0.17
0.30	0.50	0.70	0.53	0.43	[^] 0.41	0.52
0.90	0.50	0.40	0.05	[*] 0.27	[*] 0.27	0.06
0.60	0.50	0.40	0.24	[*] 0.39	[*] 0.38	0.24
0.30	0.50	0.40	0.27	0.27	0.27	0.27
0.90	0.40	0.70	0.05	[*] 0.32	[*] 0.27	0.07
0.60	0.30	0.70	0.12	0.21	0.20	0.13
0.30	0.30	0.70	0.60	[^] 0.44	[^] 0.44	0.57
0.90	0.30	0.40	0.06	[*] 0.28	[*] 0.26	0.08
0.60	0.30	0.40	0.17	[*] 0.30	[*] 0.29	0.18
0.30	0.30	0.40	0.34	0.32	0.32	0.33

FPS = 300% of baseline							
Simulated parameters							
BSL-FPS phenotypic correlation	BSL total heritability	FPS total heritability	FPS unique heritability	Estimates returned for FPS heritability		Bivariate model of raw BSL and FPS	
				Raw difference score	Standardized difference score		
0.90	0.80	0.70	0.10	[*] 0.43	0.20	0.11	
0.60	0.80	0.70	0.48	[*] 0.74	[^] 0.33	0.46	
0.30	0.80	0.70	0.67	[*] 0.81	[^] 0.51	0.64	
0.90	0.80	0.55	0.06	[*] 0.37	[*] 0.23	0.07	
0.60	0.80	0.40	0.30	[*] 0.71	0.21	0.29	
0.30	0.80	0.40	0.40	[*] 0.75	0.31	0.39	
0.90	0.50	0.70	0.12	[*] 0.45	0.21	0.13	
0.60	0.50	0.70	0.15	0.19	0.17	0.16	
0.30	0.50	0.70	0.53	0.42	[^] 0.25	0.52	
0.90	0.50	0.40	0.05	[*] 0.27	[*] 0.18	0.07	
0.60	0.50	0.40	0.24	[*] 0.39	0.22	0.24	
0.30	0.50	0.40	0.27	0.27	[^] 0.17	0.26	
0.90	0.40	0.70	0.05	[*] 0.32	[*] 0.16	0.07	
0.60	0.30	0.70	0.12	0.21	0.16	0.14	
0.30	0.30	0.70	0.60	[^] 0.44	[^] 0.24	0.57	
0.90	0.30	0.40	0.06	[*] 0.28	0.16	0.08	
0.60	0.30	0.40	0.17	[*] 0.29	0.17	0.18	
0.30	0.30	0.40	0.34	0.33	0.26	0.33	

Note: Heritability estimates returned from the univariate and bivariate models should be compared with the simulated parameter of FPS unique heritability (bolded column).

^{*} Value >10% higher than the simulated parameter

[^] Value >10% lower than the simulated parameter