International Journal of Psychophysiology 114 (2017) 24–30

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

How many blinks are necessary for a reliable startle response? A test using the NPU-threat task☆

Lynne Lieberman [a], Elizabeth S. Stevens [a], Carter J. Funkhouser [a], Anna Weinberg [b], Casey Sarapas [a], Ashley A. Huggins [c], Stewart A. Shankman [a,*]

[a] University of Illinois at Chicago, Department of Psychology, Chicago, IL 60657, United States
[b] McGill University, Department of Psychology, Montreal, QB, Canada
[c] University of Wisconsin, Department of Psychology, Milwaukee, WI 53211, United States

ARTICLE INFO

Article history:
Received 19 July 2016
Received in revised form 17 January 2017
Accepted 31 January 2017
Available online 2 February 2017

Keywords:
Anxiety-potentiated startle
Emotion-modulated startle
Eyeblink startle reflex
Fear-potentiated startle
Reliability

ABSTRACT

Emotion-modulated startle is a frequently used method in affective science. Although there is a growing literature on the reliability of this measure, it is presently unclear how many startle responses are necessary to obtain a reliable signal. The present study therefore evaluated the reliability of startle responding as a function of number of startle responses (NoS) during a widely used threat-of-shock paradigm, the NPU-threat task, in a clinical ($N = 205$) and non-clinical ($N = 92$) sample. In the clinical sample, internal consistency was also examined independently for healthy controls vs. those with panic disorder and/or major depression and retest reliability was assessed as a function of NoS. Although results varied somewhat by diagnosis and for retest reliability, the overall pattern of results suggested that six startle responses per condition were necessary to obtain acceptable reliability in clinical and non-clinical samples during this threat-of-shock paradigm in the present study.

1. Introduction

Establishing the reliability of a measure is an essential first step towards establishing its validity (Cronbach, 1947; Cronbach and Meehl, 1955). Although this fact is well accepted in the development of self-report and interview measures, the psychometric properties of psychophysiological indices of psychological constructs has received less attention until recently (Hajcak and Patrick, 2015; Tomarken, 1995). This is particularly important given the increasingly prominent role of psychophysiological measures within psychology (and affective science more specifically; Schwartz et al., 2016; Shankman and Gorka, 2015). The present study therefore seeks to contribute to this burgeoning literature by examining the reliability of a widely used psychophysiological index of emotion – electromyography of the eyeblink startle reflex (EMG startle).

The startle reflex is particularly conducive to translational research on emotion because it is present across species and its magnitude is modulated by an organism's emotional state. More specifically, the magnitude of the startle reflex is potentiated or blunted relative to baseline when an organism is in an aversive (e.g., fear) or appetitive (e.g., excitement) emotional state, respectively (Grillon and Ameli, 2001; Vrana et al., 1988). Startle is also commonly used to examine emotional processing abnormalities that may contribute to the development and maintenance of psychopathology. For example, heightened aversive responding to particular threatening stimuli/situations has been implicated in the pathogenesis of several internalizing disorders (e.g., panic disorder and interoceptive cues; posttraumatic stress disorder and trauma-related cues; social anxiety disorder and social evaluation; Craske et al., 2009). However, *unpredictable* threatening stimuli are particularly aversive for anxious individuals. Panic disorder (PD), posttraumatic stress disorder, and social anxiety disorder have all been associated with heightened startle potentiation during the anticipation of unpredictable threat (Cornwell et al., 2006; Grillon et al., 2009; Shankman et al., 2013). Thus, aberrant emotion-modulated startle, particularly during the anticipation of unpredictable threat, may represent a transdiagnostic marker for several internalizing disorders.

The literature on the psychometric properties of emotion-modulated startle has also grown in recent years. Investigations of the retest reliability of emotion-modulated startle elicited during an affective picture-viewing task have yielded mixed results, with some investigations finding strong retest reliability (Bradley et al., 1995; Larson et al., 2000) and others finding weak retest reliability (Kaye et al., 2016; Manber et al., 2000). Only two studies to date have examined retest reliability of emotion-modulated startle during the No threat-Predictable threat-Unpredictable threat-task (NPU; Schmitz and Grillon, 2012), a startle paradigm that is widely used to differentiate startle potentiation

☆ Funding: This study was supported by grants from the National Institute of Mental Health (S.S., grant numbers R01 MH098093 and R21 MH080689).

* Corresponding author at: University of Illinois at Chicago, 1007 W. Harrison St. (M/C 285), Chicago, IL 60657, United States.
E-mail address: stewarts@uic.edu (S.A. Shankman).

http://dx.doi.org/10.1016/j.ijpsycho.2017.01.012
0167-8760/

to predictable threat (i.e., fear-potentiated startle) and unpredictable threat (i.e., anxiety potentiated startle). Both studies reported retest correlations above 0.69 for anxiety-potentiated startle and fear-potentiated startle (Kaye et al., 2016; Shankman et al., 2013). Kaye et al. (2016) reported acceptable internal consistency (i.e., Cronbach's alphas > 0.70 [Nunnally, 1978]) for anxiety-potentiated startle and fear-potentiated startle during the NPU-threat task.

Despite growing focus in the field of psychology on exploring the reliability of emotion-modulated startle, there are several major gaps in the extant literature on the psychometric properties of this psychophysiological measure. For example, it is presently unknown how many startle responses are necessary to obtain a reliable index of startle potentiation scores during emotion-modulated startle paradigms. It is also presently unknown whether the number of startle responses (NoS) necessary for reliable condition averages (which are used to calculate startle potentiation scores) and potentiation scores differs for those with internalizing psychopathology relative to those without. This is a particularly important question to address given the abovementioned association between internalizing psychopathology and aberrant emotion-modulated startle.

Condition averages and potentiation scores calculated from a sufficient NoS should demonstrate acceptable internal consistency and strong retest reliability. Determining the *minimum* number of startle responses (NoS) necessary for reliable condition averages and potentiation scores would be highly beneficial for the design of future experimental protocols (at least with the NPU startle paradigm), which should be as brief as possible to reduce participant burden and the potential impact of startle habituation on task effects (Blumenthal et al., 2005). An empirically determined minimum NoS could also help experimenters determine when a participant has too few usable startle responses to be included in data analyses. This is critical given that certain trials may be excluded for some participants due to artifacts (e.g., excessive participant movement just before or after the presentation of a startle probe) and non-responses (i.e., failure to exhibit a discernable startle response) and some participants may withdraw from the study prior to study completion.

Several studies have examined the reliability of event-related potentials as a function of number of trials (e.g., Foti et al., 2013; Moran et al., 2013; Meyer et al., 2013). However, only one study to our knowledge has examined this question with respect to EMG startle data. Our laboratory recently investigated the NoS necessary for adequate internal consistency (i.e., degree of interrelatedness or stability; Tavakol and Dennick, 2011) of average startle magnitude during each condition of the NPU-threat task (i.e., condition averages) in a non-clinical sample. Startle magnitude exhibited excellent internal consistency (Cronbach's alpha > 0.80) for all NPU conditions with as few as three responses (Nelson et al., 2015). The present study will replicate our previous investigation by examining the internal consistency of condition averages during NPU as a function of NoS across *two* additional samples, one clinical and one non-clinical. We will also extend our previous investigation by examining; (a) the internal consistency of potentiation scores (i.e., fear-potentiated startle and anxiety potentiated startle) as a function of NoS; and (b) whether the NoS necessary for adequate consistency of condition averages and potentiation scores differs for those with an anxiety and/or depressive disorder. Lastly, we will conduct exploratory analyses to assess the NoS necessary for significant retest reliability of condition averages and potentiation scores in a subset of participants.

2. Methods

2.1. Participants

Data from the present study was collected as part of two investigations on emotional and cognitive processes. Details of the two studies are provided elsewhere (see Sarapas et al., 2017; Shankman et al., 2013). In brief, Study 1 ($n = 92$) was a non-clinical sample of undergraduates. Study 2 ($n = 205$) was a clinical sample recruited from the community to be in one of four groups: (1) no history of Axis I psychopathology (i.e., healthy controls; $n = 82$), (2) current major depressive disorder (MDD) and no lifetime history of any anxiety disorder (i.e., MDD-only group; $n = 37$), (3) current PD and no lifetime history of MDD (i.e., PD-only group; $n = 28$), (4) current PD and MDD (i.e., comorbid PD and MDD group; $n = 58$). Diagnoses were made via the Structured Clinical Interview for DSM-IV (SCID; First et al., 1996).

Exclusion criteria for both studies were a history of head trauma, left-handedness, and English fluency. Participants in Study 2 were additionally required to have no lifetime history of a psychotic disorder, bipolar disorder, or dementia. Participant demographics can be found in Table 1, along with clinical characteristics, such as self-reported anxiety and depressive symptomology.

2.2. Procedure and NPU-threat task

The full procedure for Studies 1 and 2 has been reported elsewhere (Sarapas et al., 2017; Shankman et al., 2013). In brief, after informed consent all participants completed the NPU threat-task. For Study 2, 34 participants returned to the laboratory 5–17 ($M = 9.46$, $SD = 3.71$) days after their initial visit to complete NPU a second time. Of these 34 individuals, 7 had MDD-only, 5 had PD-only, 10 had comorbid PD and MDD, and 12 were healthy controls. All procedures were approved by the local Institutional Review Board.

The NPU-threat task was designed to assess responses to predictable and unpredictable threats (Schmitz and Grillon, 2012). In brief, prior to the task, shock electrodes were placed on participants' left wrist and a shock work-up procedure was completed to identify the level of shock intensity each participant described as “highly annoying but not painful” (between 1 and 5 mA). Participants also completed a 2-min startle habituation task prior to the task to reduce early, exaggerated startle potentiation.

The NPU-threat task included three within-subjects conditions - no shock (N), predictable shock (P), and unpredictable shock (U). Text at the bottom of the computer monitor informed participants of the current threat condition and each condition lasted for 90 s. In Study 1, a 6-s countdown was displayed five times within each condition, and in Study 2, an 8-s geometric cue (blue circle for N, red square for P, and green star for U) was presented four times within each condition. Interstimulus intervals ranged from 7 to 17 s during which only the text describing the condition was on the screen (i.e., ISI conditions).

During N, no shocks were delivered. During P, Study 1 participants only received a shock when the countdown reached 1 and Study 2 participants only received a shock when the cue (red square) was on the screen (i.e., the shock was predicted by the countdown or cue in Studies 1 and 2, respectively). In the U condition, shocks were administered at any time (i.e., during the cue countdown [hereafter: cue] or ISI). Study 1 participants received 20 shocks (10 each during P and U) and 48 startle probes (16 each during N, P, and U). Study 2 participants received 12 shocks (6 during P and 6 during U) and 72 startle probes (24 each during N, P, and U). Study 2's NPU was divided into two recording blocks, separated by a rest period.

Table 1
Sample demographics and clinical characteristics.

Characteristic	Clinical sample	Non-clinical sample
Age	32.93 (12.31)	19.02 (1.38)
Gender (% female)	64.40	76.1
Ethnicity (% Caucasian)	46.30	35.9
IDAS-dysphoria	22.26 (10.61)	21.74 (81.90)
IDAS-panic	11.93 (5.34)	11.78 (4.00)
IUS-12	28.22 (10.09)	27.74 (8.67)

Note. IDAS = Inventory for Depression and Anxiety Symptoms (Watson et al., 2007); IUS-12 = Intolerance of Uncertainty scale (Carleton et al., 2007).