



Published in final edited form as:

*Psychophysiology*. 2016 July ; 53(7): 991–1007. doi:10.1111/psyp.12646.

## One Year Developmental Stability and Covariance among Oddball, Novelty, Go/Nogo, and Flanker Event-Related Potentials in Adolescence: a Monozygotic Twin Study

Scott J. Burwell, Stephen M. Malone, and William G. Iacono

University of Minnesota

### Abstract

Event-related potential (ERP) measures may index genetic risk for psychopathology before disorder-onset in adolescence, but little is known about their developmental rank-order stability during this period of significant brain maturation. We studied ERP stability in 48 pairs of identical twins (age 14 – 16 years) tested one year apart. Trial-averaged voltage waveforms were extracted from electroencephalographic recordings from Oddball/Novelty, Go/Nogo, and Flanker tasks, and 16 amplitude measures were examined. Members of twin pairs were highly similar, whether based on ERP amplitude measures (intraclass correlation [ICC] median = .64, range = .44 – .86) or 3 factor scores (all ICCs = .69) derived from them. Stability was high overall, with 69% of the 16 individual measures generating stability coefficients exceeding .70 and all factor scores showing stability above .75. Measures from 10 difference waveforms calculated from paired conditions within tasks were also examined, and were associated with lower twin similarity (ICC median = .52, .38 – .64) and developmental stability (only 30% exceeding .70). In a supplemental analysis, we found significant developmental stability for error-related negativity (range = .45 – .55) and positivity (.56 – .70) measures when average waveforms were based on 1 or more trials, and that these values were equivalent to those derived from averages using the current field recommendation which requires 6 or more trials. Overall, we conclude that the studied brain measures are largely stable over one year of mid- to late-adolescence, likely reflecting familial etiologic influences on brain functions pertaining to cognitive control and salience recognition.

### Keywords

event-related potential; endophenotype; developmental stability; adolescence; twin similarity

---

Individual differences in the amplitudes of several event-related potential (ERP) measures correlate with cognitive function and psychiatric disorder. Some ERP indices may serve as endophenotypes, laboratory measures that result from the same genetic liability that predispose an individual to a given disorder or spectrum of disorders. Various authors have documented the properties that endophenotypes are expected to possess (Almasy & Blangero, 2001; de Geus, 2002; Gottesman & Gould, 2003; Iacono, 1998; Iacono & Malone, 2011), but one important criterion often neglected is that endophenotypes should identify

risk for psychopathology prior to the onset of symptoms, giving them the potential both to forecast the likelihood of succumbing to disorder and to provide insight into pathophysiological processes leading to disorder development. For instance, our group (Carlson, Iacono, & McGue, 2004; Carlson, McLarnon, & Iacono, 2007; Iacono, Carlson, Malone, & McGue, 2002; Yoon, Malone, & Iacono, 2015) has found low amplitude of one ERP metric acquired at age 17 predicts subsequent development of new psychopathology up to twelve years later. Pathophysiological implications of this endophenotype suggested by others (Begleiter & Porjesz, 1999; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Polich, 2007) may eventually lead to improved diagnosis and personalized interventions. However, as Iacono and Malone (2011) noted in their paper on developmental endophenotypes, sensitivity to changing developmental influences and the age at which such endophenotypes may be profitably used to index genetic risk remains unknown. Endophenotypes that are present during fetal development and remain relatively unchanged thereafter are desirable. Yet, adolescence may impress changes upon ERP metrics during this time of substantial maturational change and emergence of psychopathology (Paus, Keshavan, & Giedd, 2008). Thus, there is considerable value in understanding how candidate endophenotypes change in adolescence and to what degree their ability to index familial risk is colored by developmental change. In this report, we address this issue by examining the developmental stability of candidate endophenotype ERP measures and determining the degree to which members of the same family (in this case, monozygotic [MZ] twins) show endophenotypic similarity during the ages of 14 to 17 years.

Several paradigms from cognitive psychology have been adapted for ERP research to probe the development and pathophysiology of neurocognitive processes. One of the oldest examples is the “Oddball task” (Squires, Squires, & Hillyard, 1975), whereby infrequently-occurring “Target” stimuli are mixed within serial presentation of frequently-occurring non-Target (“Standard”) stimuli and subjects are instructed to attend to the former while ignoring the latter. Targets elicit a larger-amplitude P3 (often referred to as P3b; see Polich, 2007 for a review) at parietal electrodes and this potentiation is thought to implicate “top-down” information processing (Nieuwenhuis et al., 2005). Some researchers have also inserted task-irrelevant “Novelty” stimuli into the Oddball task (Courchesne, Hillyard, & Galambos, 1975), which elicits an earlier-peaking and more frontally-focused “P3a” wave thought to be reflective of the brain’s “orienting response” (a bottom-up-driven attentional shift to unanticipated changes in one’s environment; Friedman, Cycowicz, & Gaeta, 2001). Amplitude reductions for both P3a (Keage et al., 2006; Rodriguez Holguin, Porjesz, Chorlian, Polich, & Begleiter, 1999a, 1999b) and P3b (Carlson, Thai, & McLarnon, 2009; Euser et al., 2012; Gao & Raine, 2009; Iacono et al., 2002; Yoon, Iacono, Malone, Bernat, & McGue, 2008) have been associated with substance use disorders and other externalizing psychopathology (e.g., attention-deficit hyperactivity disorder or ADHD, antisocial behavior). However, only P3b has gained attention as a candidate endophenotype (Euser et al., 2012; Hesselbrock, Begleiter, Porjesz, O’Connor, & Bauer, 2001; Iacono et al., 2002; McLoughlin, Makeig, & Tsuang, 2014), leaving unknown whether P3a has similar value or the degree to which it may tap into the same psychophysiological dimension.

While Oddball and Novelty stimuli elicit robust P3 waves, tasks that require substantial cognitive control over one’s behaviors (e.g., suppression of a primed response) generate a

robust negative-going N2 wave that occurs over frontal electrodes. For example, the Go/Nogo paradigm requires responding (e.g., button-press) to all frequently-occurring “Go” trials and inhibiting responses to rarely-occurring “Nogo” trials. In the Flanker task, each stimulus presented is classified as either “Congruent” (e.g., ←←←←←←←←) or “Incongruent” (→→←←→→), and correct trial performance requires responding only to the center stimulus (e.g., left-button to “←”) while ignoring distracting flanking stimuli. In both tasks, N2 is typically largest (most negative) on rarely-occurring “Nogo” (Folstein & Van Petten, 2008; Pfefferbaum, Ford, Weller, & Kopell, 1985) and “Incongruent” (Kopp, Rist, & Mattler, 1996) trials, and has been linked to theories of cognitive control and stimulus-template mismatch (Folstein & Van Petten, 2008; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004). Although little attention is paid to P3 in the Flanker task, enhanced P3 on “Nogo” trials has been posited to reflect either motor response-inhibition (Bokura, Yamaguchi, & Kobayashi, 2001; Fallgatter & Strik, 1999; Smith, Johnstone, & Barry, 2008) or Target detection/selection processes (similar to that of P3b; Kopp, Mattler, Goertz, & Rist, 1996; Salisbury, Griggs, Shenton, & McCarley, 2004). Disinhibitory psychopathology has been related to smaller amplitudes on cognitive control tasks for both N2 (Albrecht et al., 2008; McLoughlin et al., 2009) and P3 (Albrecht et al., 2013; Fallgatter et al., 2005; McLoughlin et al., 2010), but only Flanker-N2 has been posited as an endophenotype (Albrecht et al., 2008; McLoughlin et al., 2009).

The taxing nature (e.g., fast-paced, stimulus conflict) of cognitive control tasks leads to occasional performance errors. In the early 1990s, researchers observed a sharp negative voltage at frontal electrode sites shortly following the commission of erroneous responses during these tasks (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). The error-related negativity (ERN) and positivity (Pe, occurring 200 to 400 milliseconds post-error with more posterior focus) have since been investigated widely with respect to cognitive functions and psychiatric disorders. Popular theories suggest that ERN could embody an “alarm” by which future trial performance can be adapted either through reinforcement-learning (Holroyd & Coles, 2002) or conflict-monitoring (Yeung, Botvinick, & Cohen, 2004). In comparison, Pe amplitude appears to be specifically related to *awareness* of one’s mistake (Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). ERN’s correct-trial analog CRN (for correct-response negativity; Ford, 1999) is smaller in amplitude and has been viewed both in terms of its motor-generative (Ikeda et al., 1995; Pedersen et al., 1998) and behavioral-adaptive (Vidal, Burle, Bonnet, Grapperon, & Hasbroucq, 2003) roles, but is less well-researched. Enhanced and blunted ERN amplitude may provide endophenotypic value for internalizing (e.g., anxiety, depression) and externalizing disorder prediction (respectively; Euser, Evans, Greaves-Lord, Huizink, & Franken, 2013; Hajcak, Franklin, Foa, & Simons, 2008; Olvet & Hajcak, 2008), but to date, little is known about the potential of CRN and Pe to serve as endophenotypes.

The above tasks incorporate a broad sampling of neurocognitive processes (e.g., orienting, cognitive control, error-processing, etc.) often investigated in psychiatric literature and thought to be relevant to the etiology of different psychopathologies. Of these ERP measures, heritability has only extensively been studied for P3b (meta-heritability = 60%; van Beijsterveldt & van Baal, 2002). Heritability estimates for CRN, ERN, and Pe from the Flanker task (Anokhin, Golosheykin, & Heath, 2008) and Nogo-N2, Go-P3, and Nogo-P3

from a continuous performance task (Anokhin, Heath, & Myers, 2004) range about 46% to 65%, but these two studies lack replication. The degree to which measures derived from other tasks (e.g., P3a, N2 from Flanker task, ERN from Go/Nogo task) show familial resemblance remains virtually unknown. MZ twins who are reared together have more in common with each other than any other pair of family members (identical genome, shared upbringing), offering a unique opportunity for studying familial influence. In addition, because MZ twins are psychometrically equivalent to parallel forms of the same individual, high MZ twin similarity attests directly to the high reliability of a measure (Lykken, 1982).

The brain undergoes significant reorganization over the course of late childhood and adolescence (Giedd et al., 1999; Gogtay et al., 2004; Huttenlocher, 1984; Huttenlocher & Dabholkar, 1997; Huttenlocher, de Courten, Garey, & Van der Loos, 1982; Sowell et al., 2003; Sowell et al., 2004) leading to maturational changes that can reasonably be expected to influence ERP expression. Mean-levels in visually-elicited P3b amplitude decrease over adolescence (Carlson & Iacono, 2006; Courchesne, 1978; Hill et al., 1999; Katsanis, Iacono, & McGue, 1996). P3a amplitude appears to increase until mid/late adolescence, and declines thereafter (Courchesne, 1978; Cycowicz, Friedman, & Rothstein, 1996). In addition, there is evident augmentation (enhanced negativity) in mean-levels of ERN amplitude across mid/late-adolescence (Davies, Segalowitz, & Gavin, 2004; Hogan, Vargha-Khadem, Kirkham, & Baldeweg, 2005; Wiersma, van der Meere, & Roeyers, 2007) accompanied by a leveling-off in the third decade (Weinberg & Hajcak, 2011). Whether there are reliable age-related changes in N2 amplitude over adolescence remains unclear (Johnstone et al., 2007; Jonkman, Lansbergen, & Stauder, 2003; Ladouceur, Dahl, & Carter, 2007). Such age-related changes reflect mean-level instability in the population, to be expected given brain structure changes over this period. As Iacono and Malone (2011) noted, a neurobehavioral measure can still serve as an endophenotype if, despite these age effects, individual differences in the rank ordering of endophenotype trait scores are preserved with growth. Of importance is the degree to which a measure exhibits developmental *rank-order stability* indicating that an individual's position in the distribution of endophenotypic scores is preserved from one age to another.

Expression of a genetically-influenced trait need not be constant over the lifespan and thus developmental stability (typically measured by the Pearson correlation coefficient,  $r$ ) of ERP measures is of concern to researchers. It is largely unknown in youth as they enter the age of risk for the onset of psychiatric disorder. Developmental stability ( $r > .5$ ; Clayton & Larson, 2013; Helmstadter, 1964; Segalowitz, Santesso, Murphy, et al., 2010) over relatively long stretches of brain development implies that the measure could convey genetic risk for a disorder in say, adolescence, similarly to how it conveys this risk in adulthood. By contrast, a candidate endophenotype with low developmental rank-order stability may index genetic risk during a limited period in life, but lose its predictive power if examined at other time-points. Stability coefficients examined in adult populations (>18 years) over short delays (<2 months) have been substantial (most values .7 to .8) for ERP measures such as P3a and P3b (Cassidy, Robertson, & O'Connell, 2012; Debener, Kranczioch, Herrmann, & Engel, 2002), N2 (Clayton & Larson, 2013) and P3 (Fallgatter et al., 2001) from Go/Nogo and Flanker tasks, as well as ERN and Pe (Olvet & Hajcak, 2009a; Segalowitz, Santesso, Murphy, et al., 2010), but such studies provide little information regarding how stable these measures are

during adolescent brain development. Among studies including younger subjects (Carlson & Iacono, 2006; Hammerer, Li, Volkle, Muller, & Lindenberger, 2013; Meyer, Bress, & Proudfit, 2014; Segalowitz & Barnes, 1993; Segalowitz, Santesso, Murphy, et al., 2010), only a few (Carlson & Iacono, 2006; Meyer et al., 2014; Segalowitz & Barnes, 1993) have examined ERP measures' developmental stability over long intervals (e.g., years) where brain development is ongoing and likely to affect values. One study investigating 2-year test-retest stability of ERN and CRN in a pediatric sample (first tested at approximately age 11; Meyer et al., 2014) reported average stabilities of .49 and .34 respectively, much lower than what has been reported for adults (e.g.,  $r$ 's ~ .66 for both; Weinberg & Hajcak, 2011), raising the possibility that developmental rank-order stability may be poorly preserved during childhood and adolescence.

The present study examined the mean-level changes and rank-order stabilities of sixteen ERP measures obtained from Novelty (P3a), Oddball (P3b), Go/Nogo (N2, P3, ERN, CRN, Pe), and Flanker (N2, P3, ERN, CRN, Pe) paradigms over one year of mid-adolescence, a period associated with many developmental brain changes and high risk for the onset of psychopathology. Our sample of MZ twins enabled us to study the familiarity (via twin correlations) of such measures. The ideal endophenotype possesses strong familiarity and is rank-order stable over this developmental window; therefore, to the extent that these measures meet such criteria, their value as candidate endophenotypes is enhanced. This assortment of tasks and ERP measures are often studied in isolation, so little is known about the degree to which they are likely to tap into the same cognitive, neurobiological, and genetic substrates. Hence, we further extended the existing literature by examining the factor structure of the 16 measures studied here (cf. Nelson, Patrick, & Bernat, 2011), which in turn provided composite measures that could also be evaluated for their potential as endophenotypes by evaluating their developmental stability and twin similarity.

## Method

### Participants

Forty-eight adolescent monozygotic twin pairs were recruited under the auspices of the Minnesota Center for Twin and Family Research to study adolescent brain development. The sample was selected to construct three equally-sized and gender-balanced age-groups, hereinafter referred to as Cohort-14 ( $M[SD]$  age = 14.4 [.3]), Cohort-15 (15.6 [.2]), and Cohort-16 (16.5 [.3]). Two assessments were completed about one-year apart ( $M[SD]$  = .95 [.04]) (for more information on the sample, see Malone, Luciana, et al., 2014; Silverman et al., 2014; Wilson, Malone, Thomas, & Iacono, 2015).

### Tasks and electrophysiological assessment

Twin pairs were tested simultaneously in one of two dimly lit sound-attenuated rooms. Response-buttons on each armrest of the subject's chair were aligned with the subject's index finger so that bodily movement to the button-press was minimized. Video surveillance by means of infrared cameras was employed to confirm that subjects followed directions properly and remained still. Before each task, subjects underwent a brief training block to ensure they understood task directions.

**Oddball/Novelty task**—We used an extension of the rotated-heads Oddball task (Begleiter, Porjesz, Bihari, & Kissin, 1984; Bernat, Malone, Williams, Patrick, & Iacono, 2007; Hill & Steinhauer, 1993; O’Connor, Bauer, Tasman, & Hesselbrock, 1994) that included an additional Novelty stimulus-type (Nelson et al., 2011; Venables & Patrick, 2014). Each stimulus appeared on screen for 100 milliseconds. Target stimuli (16.7% of trials) were one of two stylistic oval “heads” with a triangular “nose” situated atop the oval and an oblong “ear” on either the left or right side of the oval; these were presented pseudo-randomly amidst presentation of Standard (plain oval, 66.7% of trials) stimuli. Target trials were separated into “easy/unrotated” (nose oriented up) and “difficult/rotated” (nose oriented down) conditions with equal numbers of trials per condition. Subjects were instructed to respond with a left or right index finger button-press whether the “ear” for Target trials was located on the left or right side of the head, accounting for the possible rotation of the stimulus. Additionally, Novelty stimuli (16.7% of trials) consisting of images from the International Affective Picture System (IAPS) (Lang, Bradley, & Cuthbert, 2008) were interspersed throughout presentation of Target and Standard stimuli. A set of trial-unique (i.e., each image only presented once to a subject) IAPS images were chosen to represent pleasant, unpleasant, and neutral scenes; each of the three valences comprised 5.6% of all trials. The response window for Target-stimuli button press was 1500 milliseconds and intertrial interval was randomized between one and two seconds; all Standard stimuli (the ovals and the pictures) were to be ignored. For the purposes of this study, easy/difficult-right/left Target stimuli and pleasant/unpleasant/neutral Novelty stimuli were collapsed into Target and Novelty conditions, respectively.

**Go/Nogo task**—The Go/Nogo task consisted of two stimuli that appeared in the center of the screen, one at a time (e.g., *X* or *Y*). Subjects were instructed to press a button to the presentation of every stimulus (“Go”), unless the stimulus was preceded by its exact copy (“Nogo”). For example, in the sequence [*X*, *Y*, *X*, *X*, *Y*, *X*, *Y*, *X*, *Y*, *Y*], all stimuli would be considered “Go” stimuli except for the fourth and tenth. The duration of stimulus presentation was 300 milliseconds with an inter-stimulus interval of 900 milliseconds. If the subject did not respond within 1150 milliseconds or the subject responded incorrectly, a red bar was presented to indicate poor performance. The task consisted of three blocks of trials, each using a different pair of letter stimuli (block 1: *X*-*Y*, block 2: *P*-*O*, block 3: *U*-*D*). In total, Go and Nogo trials comprised 75% and 25% of the task, respectively.

**Flanker task**—The task used here is a modified version of the Eriksen flanker task (Eriksen & Eriksen, 1974) and it is adapted from the protocol that was used in Hall, Bernat, and Patrick (2007). Stimuli for the task consisted of a five-character array of the letters *S* and *H*; four Target arrays appeared pseudo-randomly in a serial-fashion with the following frequencies: “SSSSS,” (33.3%); “HHHHH,” (33.3%); “SSHSS,” (16.7%); and “HSHHH,” (16.7%). For each array, the Target stimulus is the center character and all other characters are “flankers” (to be ignored). Subjects were told to watch the screen as stimuli were presented and indicate with the left-button if the Target stimulus was an *S* and indicate with the right-button if the Target stimulus was an *H*. EEG was recorded for three blocks of 150 trials where Target stimulus hand-mapping was alternated for each block: Block 1 (*S* = right, *H* = left), Block 2 (*S* = left, *H* = right), and Block 3 (*S* = right, *H* = left). Feedback on the

person's accuracy ("Your performance so far:  $N\%$  correct") was given at the halfway-point and end of each block.

**EEG acquisition**—Dense-array EEG (61 scalp sensors; 10/10 placement cap; 1024 Hz sample rate; passband, DC to 205 Hz) was recorded continuously while participants performed each task with the BioSemi ActiveTwo system (BioSemi, Amsterdam, Netherlands). To monitor eye movement, two additional electrodes were placed superior and inferior to the right eye and another two electrodes were placed on the right and left temples. Reference electrodes were placed on each earlobe.

**EEG processing**—Data were processed offline in MATLAB (version 7.8, Mathworks Inc.) and the EEGLAB open-source software (version 9, Delorme & Makeig, 2004) as described in Burwell, Malone, Bernat, and Iacono (2014). Continuous EEG was re-sampled at 256 Hz, highpass-filtered with a windowed sinc filter (firfilt plug-in; 0.1 Hz cutoff, Kaiser window, order of 1286), and re-referenced to the averaged potential from earlobe-situated electrodes. For the purpose of ocular correction, EEG was decomposed into independent components with the Infomax algorithm (Bell & Sejnowski, 1995). Spatial and temporal features of each subject's set of independent components were correlated with those of an expected blink component (a spatial blink template and bipolar criterion channel from vertical electrooculogram channels). Components whose squared correlation coefficients exceeded threshold derived by means of an expectation maximization algorithm (Mognon, Jovicich, Bruzzone, & Buiatti, 2010) for both spatial and temporal criteria were subtracted from the data. Ocular saccades were corrected in a similar way. Finally, data were lowpass filtered (45 Hz cutoff, Kaiser window, order of 88) and epochs were generated spanning two seconds either side of stimulus-onset.

Improbable instances of data (i.e., artifacts) were pruned from the dataset at multiple junctures. First, descriptive measures (e.g., temporal variance, maximum gradient) for each electrode and one-second time-range in the continuous EEG were computed (prior to ocular-correction and epoching). If 75% or 25% of a given electrode or time-range (respectively) was deemed a statistical outlier (exceeding 4 normalized median absolute deviations relative to the median of the distribution) (Rousseeuw & Croux, 1993), it was deleted. Segments in the ocular-corrected/epoching data were subjected to the same criteria. Prior to computing ERPs, we interpolated deleted electrodes for epochs containing more than 75% of the original data via a spherical spline method (Perrin, Pernier, Bertrand, & Echallier, 1989).

### Task performance measures

Accuracy for each task was derived as the total number of correctly-performed trials divided by the sum of correctly-performed and incorrectly-performed trials. We also calculated participants' mean reaction time (RT), defined as the average button-response latency across all correctly-performed trials. For those few subjects who reversed their hand responses for one of the blocks of the Flanker task (resulting in block-accuracy approaching zero), we ignored trials from these blocks in all analyses.

## ERP measure extraction

ERPs were calculated from the epoched EEG in MATLAB using the Psychophysiology Toolbox (<http://sourceforge.net/projects/psychophys>). The voltage at each time-point relative to stimulus- or response-onset was averaged across trials, then, the mean value from a baseline period (–200 to 0 for stimulus-locked and –500 to –300 for response-locked, in milliseconds) was subtracted from each ERP wave. To facilitate comparability with past endophenotype research using these measures (Albrecht et al., 2008; Hajcak et al., 2008; Hesselbrock et al., 2001; Iacono et al., 2002; Iacono, Malone, & McGue, 2003; McLoughlin et al., 2009), we extracted ERP measures from the raw ERP waveforms which were quantified as the average voltage within a specified window. Time-windows used to derive ERP measures are as follows: P3a (275 – 550 ms) and P3b (350 – 850 ms) from the Oddball/Novelty task; N2 (225 – 325 ms) and P3 (325 – 650 ms) and response-locked ERN/CRN (–25 – 100 ms) and Pe (220 – 440 ms) from the Go/Nogo and Flanker tasks. All stimulus-locked ERPs were computed using correctly-performed trials only while response-locked ERPs were derived from error trials (ERN, Pe) and appropriate hits (CRN) from Go/Nogo and Flanker tasks only. The number of erroneous trials going into error-related ERPs were as follows: Go/Nogo first assessment ( $M[SD, \text{range}] = 26.1 [19.1, 1 - 86]$ ); Go/Nogo second assessment (15.7 [13.1, 1 – 68]); Flanker first assessment (25.3 [25.4, 2 – 162]); Flanker second assessment (14.3 [17.8, 1 – 108]).

Two subjects (one complete pair) did not complete the Go/Nogo task and therefore had no data for that task. Two additional subjects were excluded from all analyses at intake because EEG data were contaminated with artifact. Available data for correct-related (all stimulus-locked ERPs, CRN) and error-related (ERN, Pe) waveforms varied by task and assessment (due to differences in accuracy). Out of the 96 individuals who participated in the study, 94 had available ERP measures for time 1 (T1) and 96 for T2 for the Oddball/Novelty task; Go/Nogo task: correct<sub>T1</sub> = 92, incorrect<sub>T1</sub> = 92, correct<sub>T2</sub> = 96, incorrect<sub>T2</sub> = 95; Flanker task: correct<sub>T1</sub> = 94, incorrect<sub>T1</sub> = 91, correct<sub>T2</sub> = 96, incorrect<sub>T2</sub> = 93.

Figure 1 plots grand-averaged ERP waveforms and topographical head distributions of measures. Although scores were not analyzed for a correct-related positivity component (“Pc”) which lacked a clear peak in the same time-range as Pe, scores from correctly-performed trials (within the same window as the Pe) are plotted on heads for reference. Generally, negative-going ERP measures (N2s, ERNs, and CRNs) were focused fronto-centrally (indicated by deep-blue) while positive-going ERP measures (P3s and Pe) were focused parietally (deep-red)<sup>1</sup>. Thus, based on the topographical foci of these measures and prior studies using age-comparable subjects, electrode FCZ was used for extracting N2 and ERN/CRN scores (Hogan et al., 2005; Jonkman, Sniedt, & Kemner, 2007; Ladouceur et al., 2007; Meyer et al., 2014) and electrode PZ was used for extracting P3 and Pe scores (Carlson & Iacono, 2006; Cycowicz et al., 1996; Jonkman et al., 2007; Katsanis et al., 1996; Ladouceur et al., 2007).

<sup>1</sup>Note that the posterior focus of P3a is consistent with prior research of this age-range (Courchesne, 1978; Cycowicz et al., 1996), albeit in adults P3a has a more frontal distribution (Polich, 2007).



## Statistical Analyses

We utilized linear mixed-effects regression using the nlme package (LME; Pinheiro et al., 2014) in R (R Core Team, 2013) which enabled testing for age-dependent changes (mean-level instability) in task performance and ERP measures while accounting for the nesting of repeated measures within twins nested within the same family. In the model,

$$Y_{ijk} = B_0 + B_1 \text{gender}_{jk} + B_2 \text{age}_{ijk} + \alpha_{jk} + \alpha_k + \varepsilon_{ijk},$$

$Y_{ijk}$  is the expected value of  $Y$  (accuracy, RT, or ERP score) for the  $i^{\text{th}}$  measurement (T1 or T2) of the  $j^{\text{th}}$  member of family  $k$ . The intercept, gender, and age are fixed effects, with age a time-varying covariate, and  $B_0$ ,  $B_1$ , and  $B_2$  are their associated coefficients (respectively), estimated via restricted maximum likelihood.  $B_1$  corresponds to the average effect on  $Y$  when the participant is female, and  $B_2$  the average rate of change in  $Y$  for every 1 year increase in age. In this notation,  $a_{jk}$  and  $a_k$  are random effects, which account for the correlation across repeated measures within each individual and the correlation within twin pairs, respectively. Their mean is 0 and they assumed normally distributed with an unknown variance, which is estimated simultaneously with the fixed effects.  $\varepsilon_{ijk}$  is individual-specific random error. For the sake of brevity, we only reported age-effects ( $B_2$ ).

Because two measurements are insufficient to dissociate purely age-related effects from effects that could have been influenced by prior task-exposure (i.e., “practice” or “training” effects from T1 that carry over to T2), we conducted independent samples t-tests comparing the means of those age 15 at study intake (T1 for Cohort-15) to those who were 14 at study intake but 15 when retested (T2 for Cohort-14). Similarly, those who were age 16 at study intake (Cohort-16 at T1) were compared to those who were 15 at study intake but 16 at retest (Cohort-15 at T2). To conclude that there was a significant effect of task-exposure, we required both age 15 and age 16 tests to have significant t-values ( $p$ 's < .05) indicating change in the same direction for a given measure.

We assessed developmental stability using Pearson correlation coefficients which reflect consistency in the relative ordering of individuals' ERP measures from T1 to T2. As discussed, developmental stability has implications for these measures' clinical and research utility; good stability ( $r$ 's > .5) confers “traitness” and implies that individual differences on such measures are due to stable personal factors rather than one's mental or developmental state (e.g., fatigue, mental illness, puberty) at the time of testing (Gottesman & Gould, 2003; Iacono & Malone, 2011). Confidence intervals for Pearson  $r$ 's were calculated by Fisher-transforming the coefficient, calculating its 95% interval (with  $N = 88$ , the lowest common denominator between T1 and T2  $N$ 's), and converting back into original units.

We also examined the correlational structure of our ERP measures to explore similarity among measures. Pearson correlations were calculated within-person across measures and also between members of twin pairs for each measure (in double-entry fashion, to remove effects caused by the order with which members of a twin pair were entered). The within person correlations indicate the degree to which variance is shared across measures. The within-pair/cross-measure correlations provide insight into the degree to which covariance

across measures can be attributed to familial factors (genetic similarity and similarity arising from twins being reared together). Also, for all measures in the study, we calculated intraclass twin correlations (ICCs) and accompanying 95% confidence intervals in the method of Shrout and Fleiss (1979); ICCs reflect the degree to which members of the same twin pair resemble one another on the same ERP measure (e.g., twin A's P3b to twin B's P3b). For a measure to serve as an endophenotype, these values should be high. Given the number of statistical tests, we adopted a significance threshold of  $p < .01$  for LME regression results and correlation coefficients.

These 16 ERP measures were also represented in terms of a reduced number of composite factors, extracted with principal components analysis (PCA) using the “psych” package in R (Revelle, 2014). For both N2 and P3, the Flanker task Congruent and Incongruent ERP measures correlated nearly perfectly, so for the PCA we averaged these scores resulting in “Flanker-N2” and “Flanker-P3.” Scree plots depicting singular values from T1 scores alone, T2 scores alone, T1-T2 averaged scores, and T1-T2 concatenated scores (2 rows per subject) all showed a “break” in modal explanatory variance between 3- and 4-factor solutions. Therefore we extracted 3 factors from the T1-T2 concatenated scores (standardizing within each measure) and rotated (Varimax method) the solution to get uncorrelated factors. Loadings for each measure are presented in Table 1 and can be interpreted as each measure's Pearson correlation with a given factor. The first factor (hereinafter referred to as P3-F; explaining 44% of variance) was characterized mainly by P3 and Pe measures (loading range = .52 – .85); the second factor (N2-F; explaining 29% of variance) was characterized by N2 measures (range = .73 – .84); and the third factor (RN-F; explaining 27% of variance) was characterized by response-negativities (CRNs and ERNs; range = .62 – .78).

## Results

### Does prior task-exposure affect task performance and ERP measure amplitude?

Our evaluation of prior task-exposure (i.e., practice) effects required that any observed change in task accuracy, RT, and ERP measure show a significant test result ( $p < .05$ ) for both the age 15 and age 16 analysis. None of these variables produced a significant result for both age groups, indicating that practice was unlikely to be an important contributor to change in these variables over the interval spanning T1 to T2.

### How does one year increase in age affect task performance?

As presented in the top part of Table 2, mean accuracy at both time points was high (>90%) and as  $B_2$ -values indicate, one year of age was associated with 2.20% to 2.79% improvement in accuracy across all tasks ( $p$ 's < .001). Although RT did not change for the Flanker task a year later, it did for the other tasks.

### How does one year increase in age affect ERP measure amplitude?

Illustrated in Figure 1, ERP waveforms at T2 (solid traces) were generally more negative than those at T1 (dotted traces), and this is reflected in descriptive statistics and negatively-signed age-related regression coefficients ( $B_2$ -values) in Table 2. For P3 measures (except

P3a) and the P3-F factor score, these decreases were significant ( $p$ 's < .001). No other measures produced significant effects.

### Rank-order stability of ERP measures

All individual ERP measures and factor scores were developmentally stable (see Figure 2, blue bars). Stabilities were large for N2s (median = .78, range = .75 – .88) and P3s (median = .77, range = .75 – .81) and significant ( $df = 86$ ,  $p$ 's < .001) across all tasks. Nogo-N2 was associated with the greatest stability ( $r = .88$ ), exceeding the 95% confidence-intervals (CIs) for all other measures (the closest competitor being P3a;  $r = .81$ , 95% CI = [.72, .87]). The median stability estimate for stimulus-locked measures (N2s, P3s;  $r = .77$ ) was also significantly greater than that for response-locked measures (CRN, ERN, Pe;  $r = .58$ , 95% CI = [.42, .70]). The weakest stability was associated with ERN (range = .45 – .55) which was nominally lower than CRN (range = .60 – .68) and Pe (range = .56 – .70). As expected, composite factors echoed these trends whereby P3-F and N2-F were comparably stable ( $r$ 's = .84 and .85, respectively) and that of RN-F was significantly lower ( $r = .63$ , 95% CI = [.48, .74]). Developmental stability did not differ as a function of gender nor did it differ as a function of age-cohort (with one exception for Go-N2, Cohort-14 = .66, Cohort-16 = .90,  $p = .01$ ).

We further sought to understand whether ERN and Pe, because they comprise a smaller number of trials than CRN or stimulus-locked ERPs (see **Method** for descriptive statistics), negatively biased the overall developmental stability of response-locked relative to stimulus-locked ERP measures (cf. Meyer et al., 2014; Olvet & Hajcak, 2009b; Pontifex et al., 2010). Guided by prior suggestions (Olvet & Hajcak, 2009b; Pontifex et al., 2010), we re-calculated developmental stability correlations for ERN and Pe measures by including only those subjects whose error-related ERP averages were derived from six or more trials. Applying this criterion, we observed no appreciable increase in stabilities for ERN and Pe measures (mean change in  $r = .03$ , range =  $-.03$  to  $.09$ ) but witnessed a considerable reduction in sample size (20 and 31 individuals [23%, 34%] were excluded based on this criterion for Go/Nogo and Flanker tasks, respectively). Additionally, we compared the stability of Go-CRN to stabilities of Go-N2 and Go-P3 (which have exactly the same trials included in their averages) and found that the point-estimates of Go-N2 ( $r = .81$ ) and Go-P3 ( $r = .75$ ) both exceed the CI of Go-CRN ( $r = .60$ , 95% CI = [.45, .72]). Thus, discrepancies in available trials for ERP averages is not solely responsible for response-locked measures having lower stability estimates than stimulus-locked measures.

### Familiarity of ERP measures

We also included twin ICCs (red bars) in Figure 2 for scores from individual ERP measures amplitudes and composite factors to investigate how similar the members of a twin pair were to each other (median = .64, range = .44 – .86). For the sake of simplicity, these were computed from the average T1-T2 score for each measure per participant (if data from either time-point were missing, we used the available raw score). As might be expected, the majority (75%) of the developmental stability estimates (reflecting 1-year rank-order consistency within person) for individual ERP measures were nominally larger than twin ICCs (reflecting similarity between two twin members from the same family). Yet, the

resemblance between the two patterns of correlations is striking; a given measure with large rank-order stability was often accompanied by a large ICC (and vice versa). ICCs for N2s were greatest (median = .79, range = .64 – .86), nominally higher than those for P3s (median = .68, range = .54 – .77) and significantly greater than ICCs for any of the response-locked ERP measures (median = .55, range = .44 – .64; none of which included N2s' median in their 95% CI). All composite factor ICCs were large (P3-F = .69, N2-F = .79, RN-F = .75).

### Commonality among ERP measures

Although these 16 ERP measures were obtained using different stimuli (often during different tasks), as the factor score results indicated, they share variance in common, a fact that is further fleshed out in Figure 3A which provides the correlation matrix for all the measures. In the figure, correlations of significant magnitude ( $p < .01$ ) are color-coded for their sign and strength (red = positive, blue = negative, white = non-significant). Out of 120 correlations among individual ERP measures, 58 (48%) were positive-valued and significant ( $df = 86$ ,  $p$ 's  $< .01$ ). All correlations among N2s ( $n = 6$ ; range = .57 – .95) and all correlations among P3s ( $n = 15$ ; .30 – .95) were significant and most were sizeable, albeit certain measures were apparent outliers. For example, of the P3 measures we studied, P3a was associated with the smallest inter-P3 correlations and non-overlapping range (median = .37, range = .30 – .46) relative to others' (median = .67, range = .56 – .95). This is also reflected in the P3a loading on P3-F in Table 1 ( $r = .52$ ), which possessed a 95% confidence interval barely overlapping the next-lowest loading (upper CI v. Flanker-Pe loading = .62 v. .62). Likewise, the Nogo-N2 N2-F-loading ( $r = .73$ ) was significantly weaker than that of other N2 measures (per confidence interval, 95% CI = [.65, .79]) which was consistent with the observed comparatively weak inter-N2 correlations in Figure 3A. We also observed moderately sized but significant correlations between P3 and Pe (10 of 12 correlations were significant, median = .57, range = .14 – .70) which was reflected in P3-F loading patterns for Pe ( $r$ 's  $> .6$ ). While 6 of 8 N2-ERN correlations (median = .35, range = .19 – .45) were significant, ERN correlated more strongly with CRN measures (3 of 4 were significant, median = .47, range = .27 – .48) and the RN-F composite consistent with related but separable neurobiology.

In the top 16 rows of Figure 3B we plotted within-pair/cross-measure Pearson correlations, reflecting the degree to which familial factors (e.g., genes, shared environment) contributed to similarity among different measures. Although 98 (82%) of the 120 within-subject cross-measure cells in Figure 3A were nominally greater than the corresponding within-pair/cross-measure cell in **3B** (range = -.09 – .31), we noted many symmetries. Of the 58 significant cross-measure correlations in the top 16 rows of **3A**, 21 (34%) of them were significant in **3B**. Like patterns of statistical significance from Figure 3A were seen in 3B; for example, all six N2 correlations (median = .59, range = .49 – .82) and almost all (12 of 15) P3 correlations (median = .43, range = .15 – .72) were significant.

The bottom three rows of Figure 3A and 3B display within-subject cross-measure and within-pair/cross-measure correlations among composite factors and individual ERP measures. Twenty-three of the 51 cells were significant in **3A** and 20 of those same cells were significant in **3B**, indicating that in terms of statistical significance, the bottom three

rows only differed by a mere 6% (3 of 51). As expected, correlations among factors were essentially zero (because of the orthogonal rotation) and significant correlations ( $p$ 's < .01) between composites and individual ERP measures were very similar to those in Table 1; P3/Pe scores correlated most often with P3-F, N2 scores correlated most often with N2-F, and CRN/ERN scores correlated most often with RN-F. Resemblance between cross-measure correlation patterns within-individuals (**3A**) and among co-twins (**3B**) is illustrated by absolute difference correlations in the upper right of **3B** (median = .07, range = 0 – .31) and provides further evidence that monozygotic twins are nearly parallel versions of the same individual (Lykken, 1982).

### Exploratory investigation of difference wave ERP measures

Although the literature on ERP endophenotypes has been focused on amplitude measures extracted directly from the scalp-recorded voltage at a given electrode (hereinafter referred to as “raw voltage”), also of interest is to what degree scores derived from difference waves, which contrast voltage from two task manipulations to potentially highlight task-specific ERP features (Luck, 2005), might have properties supporting their potential as endophenotypes. We computed difference waves using the following contrasts: Novelty minus Standard (P3a); Target minus Standard (P3b); Nogo minus Go (Nogo-N2, -P3); Incongruent minus Congruent (Flanker-N2, -P3); and Error minus Correct (ERN, Pe). Next, we extracted scores using identical windows used for the raw voltage measures and subjected these difference wave scores to the same developmental and twin analyses. Difference waves are plotted in Figure 4 and are remarkably similar across T1 (dotted traces) and T2 (solid traces) in terms of amplitude and shape. Mean amplitude levels for only 2 out of the 10 difference wave measures were significantly associated with age in regression analyses; the Novelty-Standard score increased with age by about 1.4 microvolts ( $p = .001$ ) while the Oddball-Standard score decreased with age by about 2.1 microvolts ( $p < .001$ ). Based on comparison of age 15 and age 16 groups, it was apparent that changes in amplitude spanning T1 to T2 were not due to prior task-exposure or practice.

Turning to Table 3, all developmental stability (median = .59, range = .26 – .78) and twin similarity (median = .52, range = .38 – .64) correlations calculated from difference waves were significant ( $df = 86$ ,  $p$ 's < .01). Compared to raw voltage ERP measures, all measures extracted from difference waves were associated with nominally lower developmental stability (difference minus raw voltage ERP measures,  $M = -.14$ , range =  $-.49$  to  $-.02$ ) and in 5 out of 10 cases (see Table 3), this reduction in stability was deemed significant because the difference wave estimate fell below the 95% confidence interval of the raw voltage estimate presented in Figure 2. In addition, an average decrease in twin similarity of .11 (range =  $-.44$  to  $.05$ ) was observed when using difference waves rather than raw voltage measures. This reduction in twin similarity was only significant for two measures (Incongruent-N2 and -P3 from the Flanker task), both of which yielded estimates that fell below the confidence interval of their respective raw voltage ICC estimates. Expectedly, using difference waves resulted in reductions in covariance among measures relative to raw voltage ERP waveforms (see Figure 5): only 9 (out of 45, 20%) of cross-measure and 6 (13%) of within-pair/cross-measure correlations remained significant. Although reliance on

difference waves may diminish redundancy across measures, a resulting cost is reduced developmental stability and familiarity.

## Discussion

In this report, we characterized age-related changes and developmental rank-order stability of sixteen ERP measures in a group of adolescent identical twins over one year of important brain development. Despite age-dependent changes in ERP measure amplitudes during this adolescent epoch (particularly for P3 measures), individuals' rank-order was preserved over a 1-year period characterized by important forms of brain development, such that stability correlations were comparable to those from prior research studying adult (>18 years) subjects over much shorter (<2 months) time windows when brain maturation is less likely to be as pronounced (Cassidy et al., 2012; Clayson & Larson, 2013; Debener et al., 2002; Fallgatter et al., 2001; Olvet & Hajcak, 2009a; Segalowitz, Santesso, Murphy, et al., 2010). We also examined the shared covariance among these measures and observed many substantial inter-measure correlations within-individuals (indicative of possibly shared neural generators) and also by twin-pair (indicative of familial influence). The commonality evident among individual ERP measures was effectively summarized by three factors extracted using PCA: 1) P3-F, accounting for 27 to 72% of the variance in P3 (P3a, P3b from Novelty/Oddball; and P3s from Go/Nogo and Flanker tasks) and Pe (Go/Nogo and Flanker tasks) measures, 2) N2-F, accounting for 53 to 71% of the variance in N2 measures from Go/Nogo and Flanker tasks, and 3) RN-F, explaining 38 to 61% of the variance in CRN and ERN amplitudes from Go/Nogo and Flanker tasks. Factors exhibited excellent developmental stability and like individual ERP measures, showed strong familial correlations. Finally, even though they have less phenotypic overlap than raw voltage ERP measures (perhaps due to greater task-specificity), difference wave ERP measures are less developmentally stable and less similar among twins.

It is of interest to understand how endophenotypes, which themselves can change with development, may usefully index familial risk during a developmental stage associated with both the onset of psychopathology and dramatic changes in brain structure and function (Iacono & Malone, 2011). In this investigation, P3 amplitude showed both reduction over one year of adolescent maturation and high rank order stability. The former observation is consistent with prior research (Carlson & Iacono, 2006; Courchesne, 1978; Hill et al., 1999; Katsanis et al., 1996), and may reflect the consequences of synaptic pruning and reorganization associated with increased developmental efficiency for information processing (Segalowitz, Santesso, & Jetha, 2010; Spear, 2003). The latter indicates that an adolescent's position in the distribution of P3 amplitude scores at one age is preserved when that individual ages by one year. This means that when using P3 amplitude as an endophenotype, in order to interpret an individual score as high or low, it must be properly age referenced. It also means that over this one year interval in mid-adolescence, the effect of maturation on P3 amplitude is about the same for everyone regardless of their P3 amplitude level at T1; thus, those with smaller P3 amplitudes at T1 tended to have smaller P3 amplitudes at T2. We can only speculate about what having small P3 amplitude relative to one's peers means in relation to brain maturation. However, it may be that adolescents possessing consistently smaller P3 amplitudes have relatively less synaptic density in

cortical regions that generate the P3 wave, perhaps due to reduced synaptogenesis earlier in life or premature synaptic pruning. Both factors would constrain the number of synaptic connections available to facilitate the development of age-appropriate cognition and adaptive behavior. Additional work would be required to address these possibilities. In the meantime, the present results support considering the measures we examined as potential endophenotypes even during a period of extensive brain reorganization.

Regardless of any observed mean-level change with age, our results show substantial one-year developmental stability and MZ twin similarity for the ERP measures examined, lending support to the construct validity of all measures as putative endophenotypes. We observed stability correlations that ranged from moderate (response-locked ERP measures) to very high (stimulus-locked ERP measures). Thus, these results enhance the potential for understudied ERP measures such as P3a, CRN, and Pe as endophenotypes. This conclusion must be tempered, however, by the fact that for many measures, mean-level changes were observed with increasing adolescent age. This combination of findings, showing that ERP measures are both age dependent and stable over a one-year interval, indicates that these measures cannot be used to index vulnerability for psychiatric disorder unless properly referenced to normative data from similarly aged adolescents; those factors that influence developmental change undoubtedly differ from those that influence individual differences in vulnerability.

To date, reduced amplitude of P3b is probably the most well established candidate endophenotype ERP measures for externalizing disorders (Euser et al., 2012; Hesselbrock et al., 2001; Iacono et al., 2002; McLoughlin et al., 2014). Therefore, one can reasonably posit that reduced amplitude of other measures that correlate strongly with P3b (e.g., Nogo-P3, Flanker-P3) are likely to be viable candidate endophenotypes for externalizing disorders because they are both stable and show familial resemblance. Psychophysiological endophenotypes might be especially suited for identifying genetic liability rather than specific genetic variants because recent work has shown that endophenotype measures derived from EEG and ERPs reflect the influence of a large number of variants each contributing very small effects (Iacono, Vaidyanathan, Vrieze, & Malone, 2014; Malone, Burwell, et al., 2014; Malone, Vaidyanathan, et al., 2014). Such polygenic traits are only useful for gene finding when examined in extremely large samples (tens of thousands individuals) that are difficult to ascertain using laboratory intensive measures such as those employed here. Nevertheless, our results suggest that these ERP measures may have utility for detecting high risk individuals prior to the expression of psychopathology, during a developmental period, adolescence, where intense study of this unique population may help provide insights into both how psychopathology develops and what brain mechanisms underlie risk.

We observed stimulus-locked P3 and response-locked Pe measures to share several similarities. It has been hypothesized that Pe is essentially a P3b-like wave that follows an erroneous button-response (rather than stimulus) (Falkenstein et al., 1991; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Leuthold & Sommer, 1999; Nieuwenhuis et al., 2001). Yet, to the best of our knowledge, only one study has examined the direct correlation between P3b and Pe ( $r = .43$ ; Cassidy et al., 2012). We too found zero-order correlations

between P3b and Pe measures of modest magnitude ( $r$ 's  $\sim .3$ ), but when each was correlated with P3-F, we observed their loadings to be sizable and remarkably similar (P3b  $r = .70$ , Nogo-Pe  $r = .72$ , Flanker-Pe  $r = .62$ ). Therefore, it seems that P3b (and other P3s) and Pe may reflect activation of similar neural circuitry, the former facilitating one's behavioral response to salient stimuli (Nieuwenhuis et al., 2005) and the latter facilitating one's (remedial) behavioral response to mistakes (Nieuwenhuis et al., 2001; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). On the other hand, P3a correlated weakly with P3-F (and other P3s/Pes individually), perhaps reflecting a dissociation between putative dopaminergic (P3a) and noradrenergic (P3b) processes (Polich, 2007).

Research on cognitive control and error-related brain processes has suggested that N2 and ERN are both generated by medial frontal cortex (including anterior cingulate cortex or ACC; Bekker, Kenemans, Hoeksma, Talsma, & Verbaten, 2005; Dehaene, Posner, & Tucker, 1994; Jonkman et al., 2007; Ladouceur et al., 2007; Miltner, Braun, & Coles, 1997; Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003; Van Veen & Carter, 2002), yet we found more differences than similarities between them. Despite several significant correlations between N2 and ERN (median  $r = .35$ ), they were comparatively weak (e.g., compared to P3-Pe median  $r = .57$ ) and in addition, readily dissociable by PCA. Research from functional neuroimaging has led to the theory that maintaining task "sets" (behavioral rules throughout a task) and adapting behavior when sets are violated (e.g., errors) represent separable top-down processes mediated by two neural systems known as the cingulo-opercular (including dorsal ACC, insula, and thalamus) and fronto-parietal networks (including frontal and parietal cortical regions), respectively (for review, see Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008), both of which undergo significant functional reorganization over adolescence (Fair et al., 2007). Perhaps N2 reflects activation of cingulo-opercular-mediated set-maintenance and CRN/ERN reflect motor-generative processes and/or brief communication between fronto-parietal and cingulo-opercular networks to convey demands for behavioral adjustment (for related discussions, see Luna, Padmanabhan, & O'Hearn, 2010; Segalowitz, Santesso, & Jetha, 2010).

## Limitations

Our study had a number of limitations. Although a central purpose of the endophenotype is to index genetic risk for psychopathology, we did not have sufficient power to examine associations between our measures and behavior disorders in our community based youth sample. This was in part because only a small portion of our community-based sample could be classified as ever having had a psychiatric disorder (*Diagnostic and Statistical Manual of Mental Disorders IV-TR*, American Psychiatric Association, 2000; ADHD = 12.5%; Oppositional Defiant Disorder = 5.2%; Conduct Disorder = 9.3%; Nicotine Dependence = 2.1%; Alcohol Abuse/Dependence = 3.1/2.1%; Illicit Drug Abuse/Dependence = 2.1/1%). Rather, our analyses highlight stability and twin similarity of these ERPs when subjects are at currently low rates of manifest disorder, but at potential risk for future disorder.

Also, although a linear model for age-related change in ERP levels is reasonable for the somewhat restricted age range of our subjects, we were unable to assess nonlinear (e.g., quadratic) change because only two measurements existed per subject. Still, we do not rule



out the possibility that ERP measures, like brain structure (Giedd et al., 1999; Gogtay et al., 2004; Huttenlocher, 1984; Huttenlocher & Dabholkar, 1997; Huttenlocher et al., 1982; Sowell et al., 2003; Sowell et al., 2004), may follow nonlinear maturation (e.g., Hill et al., 1999). More longitudinal studies are needed spanning a broader age range.

Weaker stability estimates were associated with response-locked (relative to stimulus-locked) ERP measures, a phenomenon that is not likely due to a disparity between correctly-(N2, P3, CRN measures) and incorrectly-performed (ERN, Pe measures) trials available for averaging (cf. Meyer et al., 2014; Olvet & Hajcak, 2009b; Pontifex et al., 2010). Response-locked ERP measures are presumably influenced by motor, somesthetic, premotor and/or supplementary motor brain circuits (e.g., Gevins et al., 1989), perhaps more-so than stimulus-locked measures such as P3b, whereas it is believed that temporal and parietal cortices contribute principally (e.g., Bledowski et al., 2004; Dien et al., 2003; Moores et al., 2003). Differences in stability could be due to the differential developmental stability of neural sources underlying response- versus stimulus-locked ERPs. In addition, baseline-correction of response-locked ERPs is especially prone to contamination from EEG related to stimulus-processing (Picton et al., 2000). Based on visual inspection and prior reports, we chose a baseline period believed to be sufficiently early as to not be affected by stimulus-related response-locked noise (-500 to -300 milliseconds pre-response; Meyer et al., 2014; Weinberg & Hajcak, 2011), albeit response-locked stabilities were still lower than stimulus-locked ERPs overall.

Developmental stability estimates for ERN and Pe derived using the adopted inclusion criterion based on relatively few trials (1 trials) were nonetheless sizeable and similar to those we observed using an inclusion criterion (6 trials) recommended by published research (Olvet & Hajcak, 2009b; Pontifex et al., 2010); this observation may have implications for current methodological practices. Critically, requiring six or more trials resulted in considerable sample attrition (up to 34%). Others have also excluded subjects based on this “six or more” inclusion criterion (e.g., Brooker & Buss, 2014; Kamijo et al., 2014; Perez et al., 2012; Riesel, Weinberg, Endrass, Meyer, & Hajcak, 2013; Themanson, Ball, Khatcherian, & Rosen, 2014; Themanson, Pontifex, Hillman, & McAuley, 2011; Themanson, Rosen, Pontifex, Hillman, & McAuley, 2012; Torpey, Hajcak, Kim, Kujawa, & Klein, 2012) and in some cases, the reduction in sample size has been substantial (e.g., 85 individuals, or approximately 21% of the sample in Torpey et al., 2012). Exclusion of study participants may restrict variance in key individual differences and reduce the overall applicability of a measure if it is to be used as an endophenotype. Similar to what we found in the present study, Maurer et al. (2015) concluded that acceptable ERN and Pe reliabilities may be achieved with fewer than six trials. Perhaps rather than adhering to published “rules of thumb” regarding the number of trials necessary for adequate signal-to-noise ratio for a given ERP measure (Larson, Baldwin, Good, & Fair, 2010; Leue, Klein, Lange, & Beauducel, 2013; Marco-Pallares, Cucurell, Munte, Strien, & Rodriguez-Fornells, 2011; Meyer et al., 2014; Olvet & Hajcak, 2009b; Pontifex et al., 2010; Rietdijk, Franken, & Thurik, 2014) which may be specific to study design or sample characteristics, researchers should choose inclusion criteria guided by reliability analyses of available data (cf. Maurer et al., 2015).

We acknowledge that commonality among measures may be partly due to overlap in task designs. For instance, longer intervals between relatively infrequently occurring stimuli (e.g., Target stimuli in Oddball/Novelty task, Nogo stimuli in Go/Nogo task, Incongruent stimuli in Flanker task) have been known to augment the amplitude of ERP measures such as P3 (C. J. Gonsalvez, Barry, Rushby, & Polich, 2007; C. L. Gonsalvez & Polich, 2002). Moreover, the “1-back” working memory task-style of our Go/Nogo task, while embodying the type of challenging design that recruits involves brain regions critical to cognitive control (Simmonds, Pekar, & Mostofsky, 2008), may have involved cognitive processes similar to those elicited using the oddball paradigm, thereby influencing the degree of observed covariance among P3 measures. Yet, these tasks are highly similar to those commonly used in the literature and our claims regarding developmental stability and twin similarity remain.

Exploratory investigation of difference waves indeed reduced commonality among ERP measures relative to raw voltage measures, as indicated by fewer significant and positive correlations (see Figure 5). However, we also found diminished developmental stability and twin similarity for difference wave measures which is to be expected: when two measures have much true score variance in common (as indicated by strong positive correlation, e.g., Congruent- and Incongruent-N2 from the Flanker task generates  $r = .95$ , see Figure 3), use of their difference score will decrease true score commonality with resulting diminishment of the signal-to-noise ratio and statistical reliability. Further, while use of difference waves may lessen the influence on raw scalp-recorded voltages of tissue conductivity and neural source activity that is constant across task manipulations, it may not faithfully capture the temporal-spatial mixing of neural source activity that is not shared across task manipulations. For example, to the extent that the unique neural sources recruited by one task manipulation (e.g., “Nogo”) differ from those recruited by another task manipulation (e.g., “Go”), the difference wave of these two manipulations at a given electrode may distort the temporal and spatial dynamics of the true brain-generated signal, with resultant cost in endophenotypic utility.

Finally, our decision to use a significance threshold of  $p < .01$  was not chosen based on the number of statistical tests performed, but rather was chosen as a reasoned compromise between the need to avoid Type I error while preserving sufficient sensitivity to evaluate our hypotheses. Regardless, our results depended little on the specific p-value cutoff we chose. For instance, for ERP measures for which we detected significant mean-level change in Table 2, all p-values were smaller than .0000008, suggesting that even with adjustment for multiple testing, results would not have changed. In Figures 2 and 3, we were concerned with the patterning of point-estimates, not p-values (although all developmental stability  $p$ 's  $< .000009$  and twin similarity  $p$ 's  $< .0008$  in Figure 2). If a formal multiple testing correction were applied to significance thresholds in Figure 3 (171 cells in each **3A** and **3B**), it is quite possible that nearly all cells would lose their hue, and regions of potential interest would not be easily illustrated for the reader. Thus, our fundamental conclusion, that rank-order stability of these measures persists over 1 year of adolescence despite mean-level instability, depended little on the p-value significance threshold adopted.

## Conclusion

Our results converge on the notion that multiple commonly studied ERP measures have the potential to serve as endophenotypes. Even with age-dependent changes in the mean-level of ERP amplitudes (particularly for P3 measures and P3-F factor scores), rank-order of individual amplitudes and factor scores over one year of mid-adolescence remained remarkably consistent and trait-like. Moreover, familial influence was evident in all measures and their covariance, findings that are consistent with the possibility that genetic influences play a role in the individual and common expression of ERP measures. Our results thus add to the literature by providing foundational evidence for the potential of the measures we examined to serve as candidate endophenotypes, even in asymptomatic youth. In conclusion, various N2s, P3s, and response-negativities may prove useful as endophenotypes for monitoring normal and risk-associated development of three separable neural systems, tapping set-maintenance, salience recognition, and response-monitoring processes, respectively.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This research was made possible by funding from grant number AA017314 from the National Institute of Alcohol Abuse and Alcoholism as well as grant numbers DA036216 and DA05147 from the National Institute on Drug Abuse.

## References

- Albrecht B, Brandeis D, Uebel H, Heinrich H, Mueller UC, Hasselhorn M, ... Banaschewski T. Action monitoring in boys with attention-deficit/hyperactivity disorder, their nonaffected siblings, and normal control subjects: evidence for an endophenotype. *Biol Psychiatry*. 2008; 64(7):615–625. DOI: 10.1016/j.biopsych.2007.12.016 [PubMed: 18339358]
- Albrecht B, Brandeis D, Uebel H, Valko L, Heinrich H, Drechsler R, ... Banaschewski T. Familiarity of neural preparation and response control in childhood attention deficit-hyperactivity disorder. *Psychol Med*. 2013; 43(9):1997–2011. DOI: 10.1017/S003329171200270X [PubMed: 23200032]
- Almasy L, Blangero J. Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. *Am J Med Genet*. 2001; 105(1):42–44. [PubMed: 11424994]
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4. Washington, DC: Author; 2000. text rev
- Amico F, Stauber J, Koutsouleris N, Frodl T. Anterior cingulate cortex gray matter abnormalities in adults with attention deficit hyperactivity disorder: a voxel-based morphometry study. *Psychiatry Res*. 2011; 191(1):31–35. DOI: 10.1016/j.psychres.2010.08.011 [PubMed: 21129938]
- Anokhin AP, Golosheykin S, Heath AC. Heritability of frontal brain function related to action monitoring. *Psychophysiology*. 2008; 45(4):524–534. DOI: 10.1111/j.1469-8986.2008.00664.x [PubMed: 18503487]
- Anokhin AP, Heath AC, Myers E. Genetics, prefrontal cortex, and cognitive control: a twin study of event-related brain potentials in a response inhibition task. *Neurosci Lett*. 2004; 368(3):314–318. DOI: 10.1016/j.neulet.2004.07.036 [PubMed: 15364418]
- Batty MJ, Liddle EB, Pitiot A, Toro R, Groom MJ, Scerif G, ... Hollis C. Cortical gray matter in attention-deficit/hyperactivity disorder: a structural magnetic resonance imaging study. *J Am Acad Child Adolesc Psychiatry*. 2010; 49(3):229–238. [PubMed: 20410712]

- Begleiter H, Porjesz B. What is inherited in the predisposition toward alcoholism? A proposed model. *Alcohol Clin Exp Res.* 1999; 23(7):1125–1135. [PubMed: 10443977]
- Begleiter H, Porjesz B, Bihari B, Kissin B. Event-related brain potentials in boys at risk for alcoholism. *Science.* 1984; 225(4669):1493–1496. [PubMed: 6474187]
- Bekker EM, Kenemans JL, Hoeksma MR, Talsma D, Verbaten MN. The pure electrophysiology of stopping. *Int J Psychophysiol.* 2005; 55(2):191–198. DOI: 10.1016/j.ijpsycho.2004.07.005 [PubMed: 15649550]
- Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 1995; 7(6):1129–1159. [PubMed: 7584893]
- Bernat EM, Malone SM, Williams WJ, Patrick CJ, Iacono WG. Decomposing delta, theta, and alpha time-frequency ERP activity from a visual oddball task using PCA. *Int J Psychophysiol.* 2007; 64(1):62–74. DOI: 10.1016/j.ijpsycho.2006.07.015 [PubMed: 17027110]
- Bledowski C, Prvulovic D, Hoechstetter K, Scherg M, Wibral M, Goebel R, Linden DE. Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study. *J Neurosci.* 2004; 24(42):9353–9360. DOI: 10.1523/JNEUROSCI.1897-04.2004 [PubMed: 15496671]
- Bokura H, Yamaguchi S, Kobayashi S. Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clin Neurophysiol.* 2001; 112(12):2224–2232. [PubMed: 11738192]
- Brooker RJ, Buss KA. Toddler fearfulness is linked to individual differences in error-related negativity during preschool. *Developmental Neuropsychology.* 2014; 39(1):1–8. DOI: 10.1080/87565641.2013.826661 [PubMed: 24405180]
- Burwell SJ, Malone SM, Bernat EM, Iacono WG. Does electroencephalogram phase variability account for reduced P3 brain potential in externalizing disorders? *Clin Neurophysiol.* 2014; 125(10):2007–2015. DOI: 10.1016/j.clinph.2014.02.020 [PubMed: 24656843]
- Carlson SR, Iacono WG. Heritability of P300 amplitude development from adolescence to adulthood. *Psychophysiology.* 2006; 43(5):470–480. DOI: 10.1111/j.1469-8986.2006.00450.x [PubMed: 16965609]
- Carlson SR, Iacono WG, McGue M. P300 amplitude in nonalcoholic adolescent twin pairs who become discordant for alcoholism as adults. *Psychophysiology.* 2004; 41(6):841–844. DOI: 10.1111/j.0048-5772.2004.00238.x [PubMed: 15563337]
- Carlson SR, McLarnon ME, Iacono WG. P300 amplitude, externalizing psychopathology, and earlier-versus later-onset substance-use disorder. *J Abnorm Psychol.* 2007; 116(3):565–577. DOI: 10.1037/0021-843X.116.3.565 [PubMed: 17696712]
- Carlson SR, Thai S, McLarnon ME. Visual P3 amplitude and self-reported psychopathic personality traits: frontal reduction is associated with self-centered impulsivity. *Psychophysiology.* 2009; 46(1):100–113. DOI: 10.1111/j.1469-8986.2008.00756.x [PubMed: 19055506]
- Carmona S, Vilarroya O, Bielsa A, Tremols V, Soliva JC, Rovira M, ... Bulbena A. Global and regional gray matter reductions in ADHD: a voxel-based morphometric study. *Neurosci Lett.* 2005; 389(2):88–93. DOI: 10.1016/j.neulet.2005.07.020 [PubMed: 16129560]
- Cassidy SM, Robertson IH, O'Connell RG. Retest reliability of event-related potentials: evidence from a variety of paradigms. *Psychophysiology.* 2012; 49(5):659–664. DOI: 10.1111/j.1469-8986.2011.01349.x [PubMed: 22335452]
- Clayson PE, Larson MJ. Psychometric properties of conflict monitoring and conflict adaptation indices: response time and conflict N2 event-related potentials. *Psychophysiology.* 2013; 50(12):1209–1219. DOI: 10.1111/psyp.12138 [PubMed: 23992600]
- Courchesne E. Neurophysiological correlates of cognitive development: changes in long-latency event-related potentials from childhood to adulthood. *Electroencephalogr Clin Neurophysiol.* 1978; 45(4):468–482. [PubMed: 81749]
- Courchesne E, Hillyard SA, Galambos R. Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalogr Clin Neurophysiol.* 1975; 39(2):131–143. [PubMed: 50210]
- Cycowicz YM, Friedman D, Rothstein M. An ERP developmental study of repetition priming by auditory novel stimuli. *Psychophysiology.* 1996; 33(6):680–690. [PubMed: 8961790]

- Davies PL, Segalowitz SJ, Gavin WJ. Development of response-monitoring ERPs in 7- to 25-year-olds. *Developmental Neuropsychology*. 2004; 25(3):355–376. DOI: 10.1207/s15326942dn2503\_6 [PubMed: 15148003]
- de Geus EJ. Introducing genetic psychophysiology. *Biol Psychol*. 2002; 61(1–2):1–10. [PubMed: 12385666]
- Debener S, Kranczioch C, Herrmann CS, Engel AK. Auditory novelty oddball allows reliable distinction of top-down and bottom-up processes of attention. *Int J Psychophysiol*. 2002; 46(1): 77–84. [PubMed: 12374648]
- Dehaene S, Posner MI, Tucker DM. Localization of a Neural System for Error-Detection and Compensation. *Psychol Sci*. 1994; 5(5):303–305. DOI: 10.1111/j.1467-9280.1994.tb00630.x
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*. 2004; 134(1):9–21. DOI: 10.1016/j.jneumeth.2003.10.009 [PubMed: 15102499]
- Dien J, Spencer KM, Donchin E. Localization of the event-related potential novelty response as defined by principal components analysis. *Brain Res Cogn Brain Res*. 2003; 17(3):637–650. [PubMed: 14561451]
- Dosenbach NU, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. A dual-networks architecture of top-down control. *Trends Cogn Sci*. 2008; 12(3):99–105. DOI: 10.1016/j.tics.2008.01.001 [PubMed: 18262825]
- Eriksen BA, Eriksen CW. Effects of Noise Letters Upon Identification of a Target Letter in a Nonsearch Task. *Percept Psychophys*. 1974; 16(1):143–149. DOI: 10.3758/Bf03203267
- Euser AS, Arends LR, Evans BE, Greaves-Lord K, Huizink AC, Franken IH. The P300 event-related brain potential as a neurobiological endophenotype for substance use disorders: a meta-analytic investigation. *Neurosci Biobehav Rev*. 2012; 36(1):572–603. DOI: 10.1016/j.neubiorev.2011.09.002 [PubMed: 21964481]
- Euser AS, Evans BE, Greaves-Lord K, Huizink AC, Franken IH. Diminished error-related brain activity as a promising endophenotype for substance-use disorders: evidence from high-risk offspring. *Addict Biol*. 2013; 18(6):970–984. DOI: 10.1111/adb.12002 [PubMed: 23145495]
- Fair DA, Dosenbach NU, Church JA, Cohen AL, Brahmbhatt S, Miezin FM, ... Schlaggar BL. Development of distinct control networks through segregation and integration. *Proc Natl Acad Sci U S A*. 2007; 104(33):13507–13512. DOI: 10.1073/pnas.0705843104 [PubMed: 17679691]
- Falkenstein M, Hohnsbein J, Hoormann J, Blanke L. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr Clin Neurophysiol*. 1991; 78(6):447–455. [PubMed: 1712280]
- Falkenstein M, Hoormann J, Christ S, Hohnsbein J. ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol*. 2000; 51(2–3):87–107. [PubMed: 10686361]
- Fallgatter AJ, Bartsch AJ, Strik WK, Mueller TJ, Eisenack SS, Neuhauser B, ... Herrmann MJ. Test-retest reliability of electrophysiological parameters related to cognitive motor control. *Clin Neurophysiol*. 2001; 112(1):198–204. [PubMed: 11137678]
- Fallgatter AJ, Ehlis AC, Rosler M, Strik WK, Blocher D, Herrmann MJ. Diminished prefrontal brain function in adults with psychopathology in childhood related to attention deficit hyperactivity disorder. *Psychiatry Res*. 2005; 138(2):157–169. DOI: 10.1016/j.psychres.2004.12.002 [PubMed: 15766638]
- Fallgatter AJ, Strik WK. The NoGo-anteriorization as a neurophysiological standard-index for cognitive response control. *Int J Psychophysiol*. 1999; 32(3):233–238. [PubMed: 10437634]
- Folstein JR, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*. 2008; 45(1):152–170. DOI: 10.1111/j.1469-8986.2007.00602.x [PubMed: 17850238]
- Ford JM. Schizophrenia: the broken P300 and beyond. *Psychophysiology*. 1999; 36(6):667–682. [PubMed: 10554581]
- Friedman D, Cycowicz YM, Gaeta H. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci Biobehav Rev*. 2001; 25(4):355–373. [PubMed: 11445140]

- Gao Y, Raine A. P3 event-related potential impairments in antisocial and psychopathic individuals: a meta-analysis. *Biol Psychol.* 2009; 82(3):199–210. DOI: 10.1016/j.biopsycho.2009.06.006 [PubMed: 19576948]
- Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E. A Neural System for Error-Detection and Compensation. *Psychol Sci.* 1993; 4(6):385–390. DOI: 10.1111/j.1467-9280.1993.tb00586.x
- Gevens AS, Bressler SL, Morgan NH, Cutillo BA, White RM, Greer DS, Illes J. Event-related covariances during a bimanual visuomotor task. I. Methods and analysis of stimulus- and response-locked data. *Electroencephalogr Clin Neurophysiol.* 1989; 74(1):58–75. [PubMed: 2463150]
- Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, ... Rapoport JL. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci.* 1999; 2(10):861–863. DOI: 10.1038/13158 [PubMed: 10491603]
- Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, ... Thompson PM. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A.* 2004; 101(21):8174–8179. DOI: 10.1073/pnas.0402680101 [PubMed: 15148381]
- Gonsalvez CJ, Barry RJ, Rushby JA, Polich J. Target-to-target interval, intensity, and P300 from an auditory single-stimulus task. *Psychophysiology.* 2007; 44(2):245–250. DOI: 10.1111/j.1469-8986.2007.00495.x [PubMed: 17343708]
- Gonsalvez CL, Polich J. P300 amplitude is determined by target-to-target interval. *Psychophysiology.* 2002; 39(3):388–396. [PubMed: 12212658]
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry.* 2003; 160(4):636–645. DOI: 10.1176/appi.ajp.160.4.636 [PubMed: 12668349]
- Hajcak G, Franklin ME, Foa EB, Simons RF. Increased error-related brain activity in pediatric obsessive-compulsive disorder before and after treatment. *Am J Psychiatry.* 2008; 165(1):116–123. DOI: 10.1176/appi.ajp.2007.07010143 [PubMed: 17986681]
- Hall JR, Bernat EM, Patrick CJ. Externalizing psychopathology and the error-related negativity. *Psychol Sci.* 2007; 18(4):326–333. DOI: 10.1111/j.1467-9280.2007.01899.x [PubMed: 17470258]
- Hammerer D, Li SC, Volkle M, Muller V, Lindenberger U. A lifespan comparison of the reliability, test-retest stability, and signal-to-noise ratio of event-related potentials assessed during performance monitoring. *Psychophysiology.* 2013; 50(1):111–123. DOI: 10.1111/j.1469-8986.2012.01476.x [PubMed: 23110313]
- Helmstadter, GC. Principles of psychological measurement. New York: Appleton-Century-Crofts; 1964.
- Hesselbrock V, Begleiter H, Porjesz B, O'Connor S, Bauer L. P300 event-related potential amplitude as an endophenotype of alcoholism—evidence from the collaborative study on the genetics of alcoholism. *J Biomed Sci.* 2001; 8(1):77–82. 54016. [PubMed: 11173979]
- Hill SY, Shen S, Locke J, Steinhauer SR, Konicky C, Lowers L, Connolly J. Developmental delay in P300 production in children at high risk for developing alcohol-related disorders. *Biol Psychiatry.* 1999; 46(7):970–981. [PubMed: 10509180]
- Hill SY, Steinhauer SR. Assessment of prepubertal and postpubertal boys and girls at risk for developing alcoholism with P300 from a visual discrimination task. *J Stud Alcohol.* 1993; 54(3):350–358. [PubMed: 8487544]
- Hogan AM, Vargha-Khadem F, Kirkham FJ, Baldeweg T. Maturation of action monitoring from adolescence to adulthood: an ERP study. *Dev Sci.* 2005; 8(6):525–534. DOI: 10.1111/j.1467-7687.2005.00444.x [PubMed: 16246244]
- Holroyd CB, Coles MG. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev.* 2002; 109(4):679–709. [PubMed: 12374324]
- Huttenlocher PR. Synapse elimination and plasticity in developing human cerebral cortex. *Am J Ment Defic.* 1984; 88(5):488–496. [PubMed: 6731486]
- Huttenlocher PR, Dabholkar AS. Regional differences in synaptogenesis in human cerebral cortex. *J Comp Neurol.* 1997; 387(2):167–178. [PubMed: 9336221]

- Huttenlocher PR, de Courten C, Garey LJ, Van der Loos H. Synaptogenesis in human visual cortex--evidence for synapse elimination during normal development. *Neurosci Lett.* 1982; 33(3):247–252. [PubMed: 7162689]
- Iacono WG. Identifying psychophysiological risk for psychopathology: examples from substance abuse and schizophrenia research. *Psychophysiology.* 1998; 35(6):621–637. [PubMed: 9844425]
- Iacono WG, Carlson SR, Malone SM, McGue M. P3 event-related potential amplitude and the risk for disinhibitory disorders in adolescent boys. *Arch Gen Psychiatry.* 2002; 59(8):750–757. [PubMed: 12150652]
- Iacono WG, Malone SM. Developmental Endophenotypes: Indexing Genetic Risk for Substance Abuse with the P300 Brain Event-Related Potential. *Child Dev Perspect.* 2011; 5(4):239–247. DOI: 10.1111/j.1750-8606.2011.00205.x [PubMed: 22247735]
- Iacono WG, Malone SM, McGue M. Substance use disorders, externalizing psychopathology, and P300 event-related potential amplitude. *Int J Psychophysiol.* 2003; 48(2):147–178. [PubMed: 12763572]
- Iacono WG, Vaidyanathan U, Vrieze SI, Malone SM. Knowns and unknowns for psychophysiological endophenotypes: integration and response to commentaries. *Psychophysiology.* 2014; 51(12): 1339–1347. DOI: 10.1111/psyp.12358 [PubMed: 25387720]
- Ikeda A, Luders HO, Shibasaki H, Collura TF, Burgess RC, Morris HH 3rd, Hamano T. Movement-related potentials associated with bilateral simultaneous and unilateral movements recorded from human supplementary motor area. *Electroencephalogr Clin Neurophysiol.* 1995; 95(5):323–334. [PubMed: 7489661]
- Johnstone SJ, Dimoska A, Smith JL, Barry RJ, Pleffer CB, Chiswick D, Clarke AR. The development of stop-signal and Go/Nogo response inhibition in children aged 7–12 years: performance and event-related potential indices. *Int J Psychophysiol.* 2007; 63(1):25–38. DOI: 10.1016/j.ijpsycho.2006.07.001 [PubMed: 16919346]
- Jonkman LM, Lansbergen M, Stauder JE. Developmental differences in behavioral and event-related brain responses associated with response preparation and inhibition in a go/nogo task. *Psychophysiology.* 2003; 40(5):752–761. [PubMed: 14696728]
- Jonkman LM, Sniedt FL, Kemner C. Source localization of the Nogo-N2: a developmental study. *Clin Neurophysiol.* 2007; 118(5):1069–1077. DOI: 10.1016/j.clinph.2007.01.017 [PubMed: 17368096]
- Kamijo K, Pontifex MB, Khan NA, Raine LB, Scudder MR, Drollette ES, ... Hillman CH. The negative association of childhood obesity to cognitive control of action monitoring. *Cereb Cortex.* 2014; 24(3):654–662. DOI: 10.1093/cercor/bhs349 [PubMed: 23146965]
- Katsanis J, Iacono WG, McGue MK. The association between P300 and age from preadolescence to early adulthood. *Int J Psychophysiol.* 1996; 24(3):213–221. [PubMed: 8993996]
- Keage HA, Clark CR, Hermens DF, Kohn MR, Clarke S, Williams LM, ... Gordon E. Distractibility in AD/HD predominantly inattentive and combined subtypes: the P3a ERP component, heart rate and performance. *J Integr Neurosci.* 2006; 5(1):139–158. [PubMed: 16544371]
- Kopp B, Mattler U, Goertz R, Rist F. N2, P3 and the lateralized readiness potential in a nogo task involving selective response priming. *Electroencephalogr Clin Neurophysiol.* 1996; 99(1):19–27. [PubMed: 8758967]
- Kopp B, Rist F, Mattler U. N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology.* 1996; 33(3):282–294. [PubMed: 8936397]
- Ladouceur CD, Dahl RE, Carter CS. Development of action monitoring through adolescence into adulthood: ERP and source localization. *Dev Sci.* 2007; 10(6):874–891. DOI: 10.1111/j.1467-7687.2007.00639.x [PubMed: 17973802]
- Lang, PJ.; Bradley, MM.; Cuthbert, BN. Technical Report A-8. University of Florida; Gainesville, FL: 2008. International affective picture system (IAPS): Affective ratings of pictures and instruction manual.
- Larson MJ, Baldwin SA, Good DA, Fair JE. Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): the role of number of trials. *Psychophysiology.* 2010; 47(6):1167–1171. DOI: 10.1111/j.1469-8986.2010.01022.x [PubMed: 20477982]

- Leue A, Klein C, Lange S, Beauducel A. Inter-individual and intra-individual variability of the N2 component: on reliability and signal-to-noise ratio. *Brain Cogn.* 2013; 83(1):61–71. DOI: 10.1016/j.bandc.2013.06.009 [PubMed: 23911659]
- Leuthold H, Sommer W. ERP correlates of error processing in spatial S-R compatibility tasks. *Clin Neurophysiol.* 1999; 110(2):342–357. [PubMed: 10210624]
- Luck, SJ. Ten simple rules for designing and interpreting ERP experiments. In: Handy, TC., editor. *Event-related potentials: a methods handbook.* MIT Press; 2005.
- Luna B, Padmanabhan A, O’Hearn K. What has fMRI told us about the development of cognitive control through adolescence? *Brain Cogn.* 2010; 72(1):101–113. DOI: 10.1016/j.bandc.2009.08.005 [PubMed: 19765880]
- Lykken DT. Presidential address, 1981. Research with twins: the concept of emergence. *Psychophysiology.* 1982; 19(4):361–373. [PubMed: 6889750]
- Malone SM, Burwell SJ, Vaidyanathan U, Miller MB, McGue M, Iacono WG. Heritability and molecular-genetic basis of resting EEG activity: A genome-wide association study. *Psychophysiology.* 2014; 51(12):1225–1245. DOI: 10.1111/psyp.12344 [PubMed: 25387704]
- Malone SM, Luciana M, Wilson S, Sparks JC, Hunt RH, Thomas KM, Iacono WG. Adolescent drinking and motivated decision-making: a cotwin-control investigation with monozygotic twins. *Behav Genet.* 2014; 44(4):407–418. DOI: 10.1007/s10519-014-9651-0 [PubMed: 24676464]
- Malone SM, Vaidyanathan U, Basu S, Miller MB, McGue M, Iacono WG. Heritability and molecular-genetic basis of the P3 event-related brain potential: A genome-wide association study. *Psychophysiology.* 2014; 51(12):1246–1258. DOI: 10.1111/psyp.12345 [PubMed: 25387705]
- Marco-Pallares J, Cucurell D, Munte TF, Strien N, Rodriguez-Fornells A. On the number of trials needed for a stable feedback-related negativity. *Psychophysiology.* 2011; 48(6):852–860. DOI: 10.1111/j.1469-8986.2010.01152.x [PubMed: 21091959]
- Maurer JM, Steele VR, Edwards BG, Bernat EM, Calhoun VD, Kiehl KA. Dysfunctional error-related processing in female psychopathy. *Soc Cogn Affect Neurosci.* 2015; doi: 10.1093/scan/nsv070
- McLoughlin G, Albrecht B, Banaschewski T, Rothenberger A, Brandeis D, Asherson P, Kuntsi J. Performance monitoring is altered in adult ADHD: a familial event-related potential investigation. *Neuropsychologia.* 2009; 47(14):3134–3142. DOI: 10.1016/j.neuropsychologia.2009.07.013 [PubMed: 19643116]
- McLoughlin G, Albrecht B, Banaschewski T, Rothenberger A, Brandeis D, Asherson P, Kuntsi J. Electrophysiological evidence for abnormal preparatory states and inhibitory processing in adult ADHD. *Behav Brain Funct.* 2010; 6:66.doi: 10.1186/1744-9081-6-66 [PubMed: 21029446]
- McLoughlin G, Makeig S, Tsuang MT. In search of biomarkers in psychiatry: EEG-based measures of brain function. *Am J Med Genet B Neuropsychiatr Genet.* 2014; 165B(2):111–121. DOI: 10.1002/ajmg.b.32208 [PubMed: 24273134]
- Meyer A, Bress JN, Proudfit GH. Psychometric properties of the error-related negativity in children and adolescents. *Psychophysiology.* 2014; 51(7):602–610. DOI: 10.1111/psyp.12208 [PubMed: 24646380]
- Miltner WH, Braun CH, Coles MG. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. *J Cogn Neurosci.* 1997; 9(6):788–798. DOI: 10.1162/jocn.1997.9.6.788 [PubMed: 23964600]
- Mognon A, Jovicich J, Bruzzone L, Buiatti M. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology.* 2010; doi: 10.1111/j.1469-8986.2010.01061.x
- Moore KA, Clark CR, Hadfield JL, Brown GC, Taylor DJ, Fitzgibbon SP, ... Greenblatt R. Investigating the generators of the scalp recorded visuo-verbal P300 using cortically constrained source localization. *Hum Brain Mapp.* 2003; 18(1):53–77. DOI: 10.1002/hbm.10073 [PubMed: 12454912]
- Nelson LD, Patrick CJ, Bernat EM. Operationalizing proneness to externalizing psychopathology as a multivariate psychophysiological phenotype. *Psychophysiology.* 2011; 48(1):64–72. DOI: 10.1111/j.1469-8986.2010.01047.x [PubMed: 20573054]



- Nieuwenhuis S, Aston-Jones G, Cohen JD. Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol Bull.* 2005; 131(4):510–532. DOI: 10.1037/0033-2909.131.4.510 [PubMed: 16060800]
- Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GP, Kok A. Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology.* 2001; 38(5):752–760. [PubMed: 11577898]
- Nieuwenhuis S, Yeung N, van den Wildenberg W, Ridderinkhof KR. Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cogn Affect Behav Neurosci.* 2003; 3(1):17–26. [PubMed: 12822595]
- O'Connor S, Bauer L, Tasman A, Hesselbrock V. Reduced P3 amplitudes are associated with both a family history of alcoholism and antisocial personality disorder. *Prog Neuropsychopharmacol Biol Psychiatry.* 1994; 18(8):1307–1321. [PubMed: 7863018]
- Olivet DM, Hajcak G. The error-related negativity (ERN) and psychopathology: toward an endophenotype. *Clin Psychol Rev.* 2008; 28(8):1343–1354. DOI: 10.1016/j.cpr.2008.07.003 [PubMed: 18694617]
- Olivet DM, Hajcak G. Reliability of error-related brain activity. *Brain Res.* 2009a; 1284:89–99. DOI: 10.1016/j.brainres.2009.05.079 [PubMed: 19501071]
- Olivet DM, Hajcak G. The stability of error-related brain activity with increasing trials. *Psychophysiology.* 2009b; 46(5):957–961. DOI: 10.1111/j.1469-8986.2009.00848.x [PubMed: 19558398]
- Overbeek T, Nieuwenhuis S, Ridderinkhof KR. Dissociable components of error processing: on the functional significance of the Pe vis-a-vis the ERN/Ne. *Journal of Psychophysiology.* 2005; 19(4):319–329. DOI: 10.1027/0269-8803.19.4.319
- Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci.* 2008; 9(12):947–957. DOI: 10.1038/nrn2513 [PubMed: 19002191]
- Pedersen JR, Johannsen P, Bak CK, Kofoed B, Saermark K, Gjedde A. Origin of human motor readiness field linked to left middle frontal gyrus by MEG and PET. *Neuroimage.* 1998; 8(2):214–220. DOI: 10.1006/nimg.1998.0362 [PubMed: 9740763]
- Perez VB, Ford JM, Roach BJ, Woods SW, McGlashan TH, Srihari VH, ... Mathalon DH. Error monitoring dysfunction across the illness course of schizophrenia. *J Abnorm Psychol.* 2012; 121(2):372–387. DOI: 10.1037/a0025487 [PubMed: 22060947]
- Perrin F, Pernier J, Bertrand O, Echallier JF. Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol.* 1989; 72(2):184–187. [PubMed: 2464490]
- Pfefferbaum A, Ford JM, Weller BJ, Kopell BS. ERPs to response production and inhibition. *Electroencephalogr Clin Neurophysiol.* 1985; 60(5):423–434. [PubMed: 2580694]
- Picton TW, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R Jr, Taylor MJ. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology.* 2000; 37(2):127–152. [PubMed: 10731765]
- Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D. the R Development Core Team. R package version 3.1–111. 2013. nlme: Linear and Nonlinear Mixed Effects Models.
- Polich J. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol.* 2007; 118(10):2128–2148. DOI: 10.1016/j.clinph.2007.04.019 [PubMed: 17573239]
- Pontifex MB, Scudder MR, Brown ML, O'Leary KC, Wu CT, Themanson JR, Hillman CH. On the number of trials necessary for stabilization of error-related brain activity across the life span. *Psychophysiology.* 2010; 47(4):767–773. DOI: 10.1111/j.1469-8986.2010.00974.x [PubMed: 20230502]
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.
- Revelle, W. psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston; Illinois, USA: 2014. <http://CRAN.R-project.org/package=psych> Version = 1.4.5
- Proal E, Reiss PT, Klein RG, Mannuzza S, Gotimer K, Ramos-Olazagasti MA, ... Castellanos FX. Brain gray matter deficits at 33-year follow-up in adults with attention-deficit/hyperactivity disorder established in childhood. *Arch Gen Psychiatry.* 2011; 68(11):1122–1134. DOI: 10.1001/archgenpsychiatry.2011.117 [PubMed: 22065528]

- Ridderinkhof KR, van den Wildenberg WP, Segalowitz SJ, Carter CS. Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain Cogn.* 2004; 56(2):129–140. DOI: 10.1016/j.bandc.2004.09.016 [PubMed: 15518930]
- Riesel A, Weinberg A, Endrass T, Meyer A, Hajcak G. The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biol Psychol.* 2013; 93(3):377–385. DOI: 10.1016/j.biopsycho.2013.04.007 [PubMed: 23607999]
- Rietdijk WJ, Franken IH, Thurik AR. Internal consistency of event-related potentials associated with cognitive control: N2/P3 and ERN/Pe. *PLoS One.* 2014; 9(7):e102672.doi: 10.1371/journal.pone.0102672 [PubMed: 25033272]
- Rodriguez Holguin S, Porjesz B, Chorlian DB, Polich J, Begleiter H. Visual P3a in male alcoholics and controls. *Alcohol Clin Exp Res.* 1999a; 23(4):582–591. [PubMed: 10235292]
- Rodriguez Holguin S, Porjesz B, Chorlian DB, Polich J, Begleiter H. Visual P3a in male subjects at high risk for alcoholism. *Biol Psychiatry.* 1999b; 46(2):281–291. [PubMed: 10418704]
- Rousseeuw PJ, Croux C. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association.* 1993; 88(424):1273–1283.
- Salisbury DF, Griggs CB, Shenton ME, McCarley RW. The NoGo P300 ‘anteriorization’ effect and response inhibition. *Clin Neurophysiol.* 2004; 115(7):1550–1558. DOI: 10.1016/j.clinph.2004.01.028 [PubMed: 15203056]
- Segalowitz SJ, Barnes KL. The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology.* 1993; 30(5):451–459. [PubMed: 8416071]
- Segalowitz SJ, Santesso DL, Jetha MK. Electrophysiological changes during adolescence: a review. *Brain Cogn.* 2010; 72(1):86–100. DOI: 10.1016/j.bandc.2009.10.003 [PubMed: 19914761]
- Segalowitz SJ, Santesso DL, Murphy TI, Homan D, Chantziantoniou DK, Khan S. Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology.* 2010; 47(2):260–270. DOI: 10.1111/j.1469-8986.2009.00942.x [PubMed: 20030755]
- Seidman LJ, Biederman J, Liang L, Valera EM, Monuteaux MC, Brown A, ... Makris N. Gray matter alterations in adults with attention-deficit/hyperactivity disorder identified by voxel based morphometry. *Biol Psychiatry.* 2011; 69(9):857–866. DOI: 10.1016/j.biopsycho.2010.09.053 [PubMed: 21183160]
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979; 86(2):420–428. [PubMed: 18839484]
- Silverman MH, Krueger RF, Iacono WG, Malone SM, Hunt RH, Thomas KM. Quantifying familial influences on brain activation during the monetary incentive delay task: An adolescent monozygotic twin study. *Biol Psychol.* 2014; 103C:7–14. DOI: 10.1016/j.biopsycho.2014.07.016 [PubMed: 25101864]
- Simmonds DJ, Pekar JJ, Mostofsky SH. Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia.* 2008; 46(1): 224–232. DOI: 10.1016/j.neuropsychologia.2007.07.015 [PubMed: 17850833]
- Smith JL, Johnstone SJ, Barry RJ. Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. *Clin Neurophysiol.* 2008; 119(3):704–714. DOI: 10.1016/j.clinph.2007.11.042 [PubMed: 18164657]
- Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across the human life span. *Nat Neurosci.* 2003; 6(3):309–315. DOI: 10.1038/nn1008 [PubMed: 12548289]
- Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW. Longitudinal mapping of cortical thickness and brain growth in normal children. *J Neurosci.* 2004; 24(38):8223–8231. DOI: 10.1523/JNEUROSCI.1798-04.2004 [PubMed: 15385605]
- Spear, LP. Neurodevelopment during adolescence. In: Cicchetti, D.; Walker, EF., editors. *Neurodevelopmental mechanisms in psychopathology.* Cambridge, UK: Cambridge; 2003. p. 62-83.
- Squires NK, Squires KC, Hillyard SA. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol.* 1975; 38(4):387–401. [PubMed: 46819]

- Themanson JR, Ball AB, Khatcherian SM, Rosen PJ. The effects of social exclusion on the ERN and the cognitive control of action monitoring. *Psychophysiology*. 2014; 51(3):215–225. [PubMed: 25003166]
- Themanson JR, Pontifex MB, Hillman CH, McAuley E. The relation of self-efficacy and error-related self-regulation. *Int J Psychophysiol*. 2011; 80(1):1–10. DOI: 10.1016/j.ijpsycho.2011.01.005 [PubMed: 21256167]
- Themanson JR, Rosen PJ, Pontifex MB, Hillman CH, McAuley E. Alterations in error-related brain activity and post-error behavior over time. *Brain Cogn*. 2012; 80(2):257–265. DOI: 10.1016/j.bandc.2012.07.003 [PubMed: 22940400]
- Torpey DC, Hajcak G, Kim J, Kujawa A, Klein DN. Electrocortical and behavioral measures of response monitoring in young children during a Go/No-Go task. *Dev Psychobiol*. 2012; 54(2): 139–150. DOI: 10.1002/dev.20590 [PubMed: 21815136]
- van Beijsterveldt CE, van Baal GC. Twin and family studies of the human electroencephalogram: a review and a meta-analysis. *Biol Psychol*. 2002; 61(1–2):111–138. [PubMed: 12385672]
- Van Veen V, Carter CS. The timing of action-monitoring processes in the anterior cingulate cortex. *J Cogn Neurosci*. 2002; 14(4):593–602. DOI: 10.1162/08989290260045837 [PubMed: 12126500]
- Venables NC, Patrick CJ. Reconciling discrepant findings for P3 brain response in criminal psychopathy through reference to the concept of externalizing proneness. *Psychophysiology*. 2014; 51(5):427–436. DOI: 10.1111/psyp.12189 [PubMed: 24579849]
- Vidal F, Burle B, Bonnet M, Grapperon J, Hasbroucq T. Error negativity on correct trials: a reexamination of available data. *Biol Psychol*. 2003; 64(3):265–282. [PubMed: 14630407]
- Weinberg A, Hajcak G. Longer term test-retest reliability of error-related brain activity. *Psychophysiology*. 2011; 48(10):1420–1425. DOI: 10.1111/j.1469-8986.2011.01206.x [PubMed: 21496055]
- Wiersema JR, van der Meere JJ, Roeyers H. Developmental changes in error monitoring: an event-related potential study. *Neuropsychologia*. 2007; 45(8):1649–1657. DOI: 10.1016/j.neuropsychologia.2007.01.004 [PubMed: 17303199]
- Wilson S, Malone SM, Thomas KM, Iacono WG. Adolescent drinking and brain morphometry: A co-twin control analysis. *Dev Cogn Neurosci*. 2015; doi: 10.1016/j.dcn.2015.07.005
- Yeung N, Botvinick MM, Cohen JD. The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol Rev*. 2004; 111(4):931–959. DOI: 10.1037/0033-295X.111.4.939 [PubMed: 15482068]
- Yoon HH, Iacono WG, Malone SM, Bernat EM, McGue M. The effects of childhood disruptive disorder comorbidity on P3 event-related brain potentials in preadolescents with ADHD. *Biol Psychol*. 2008; 79(3):329–336. DOI: 10.1016/j.biopsycho.2008.08.001 [PubMed: 18762228]
- Yoon HH, Malone SM, Iacono WG. Longitudinal stability and predictive utility of the visual P3 response in adults with externalizing psychopathology. *Psychophysiology*. 2015; doi: 10.1111/psyp.12548

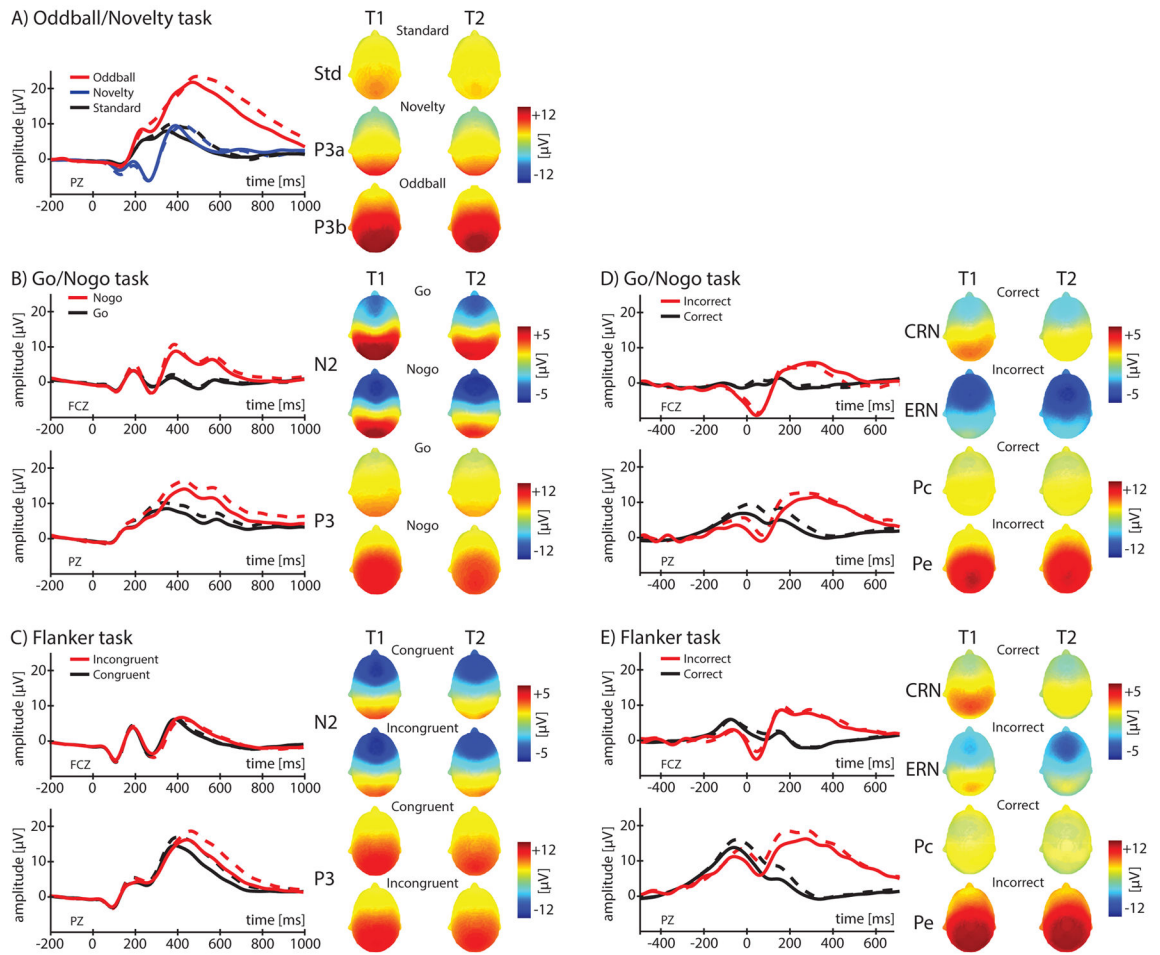


Figure 1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

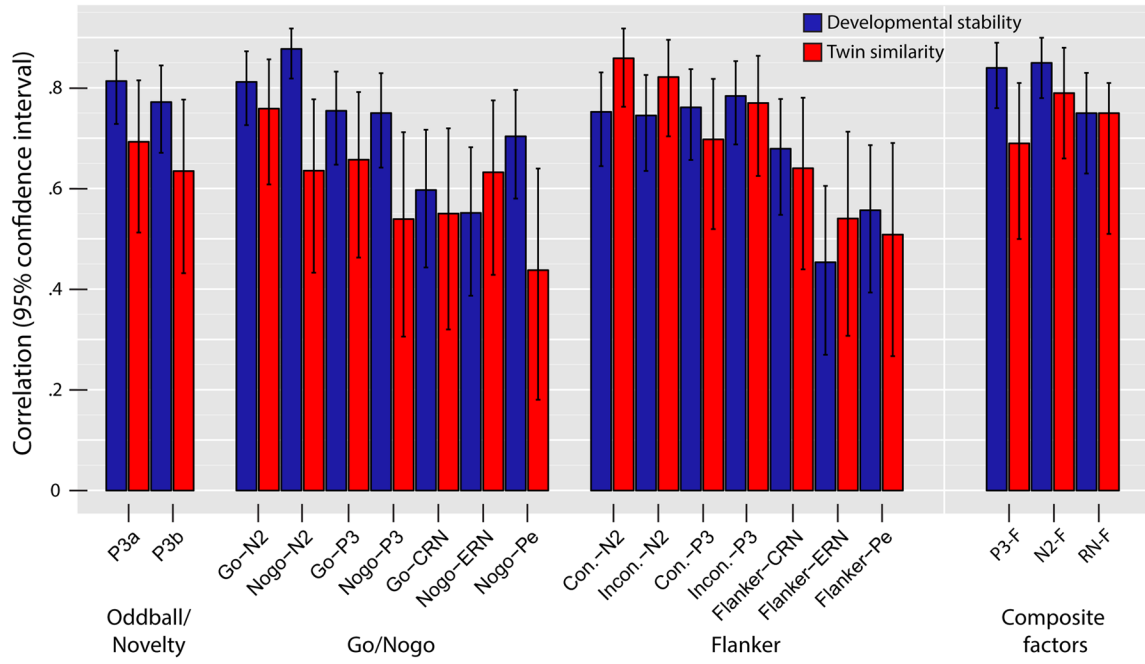


Figure 2.

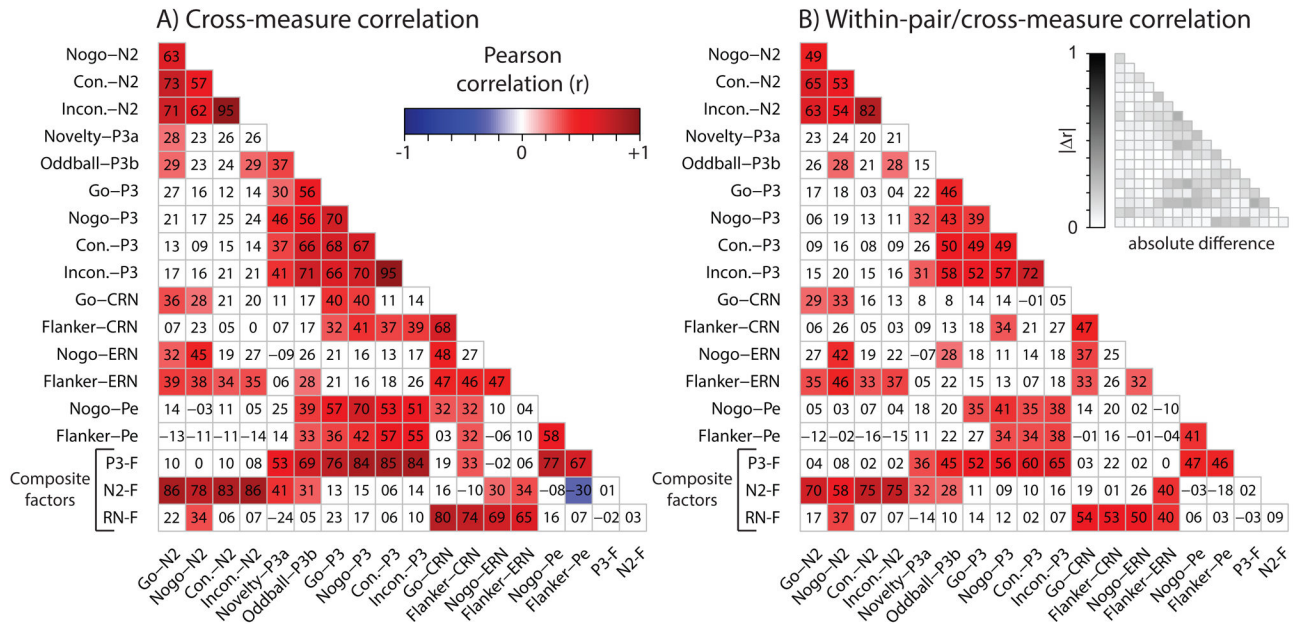
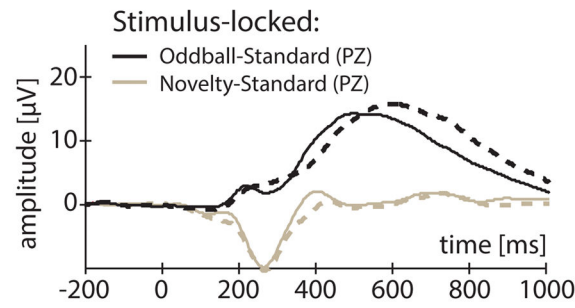
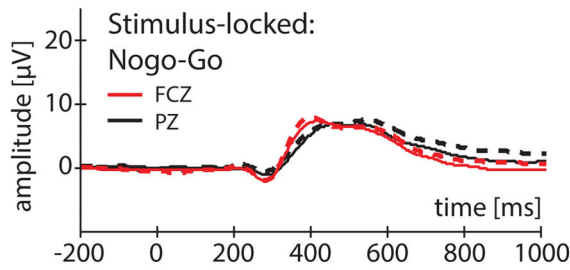


Figure 3.

A) Oddball/Novelty task



B) Go/Nogo task



C) Flanker task

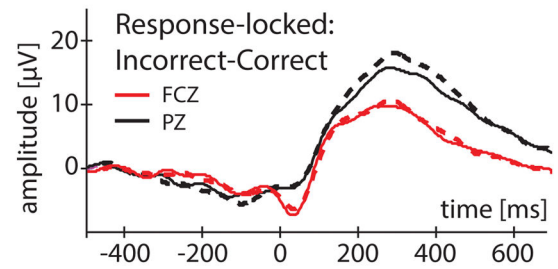
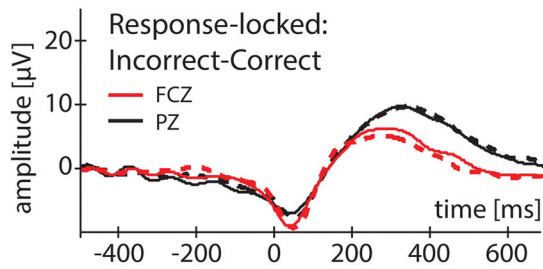
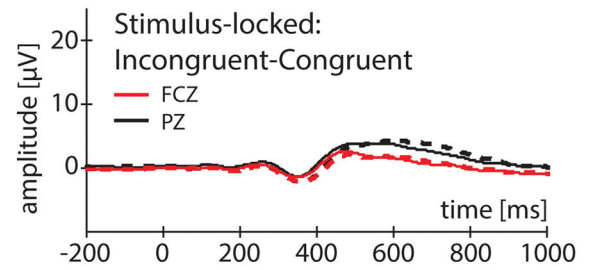
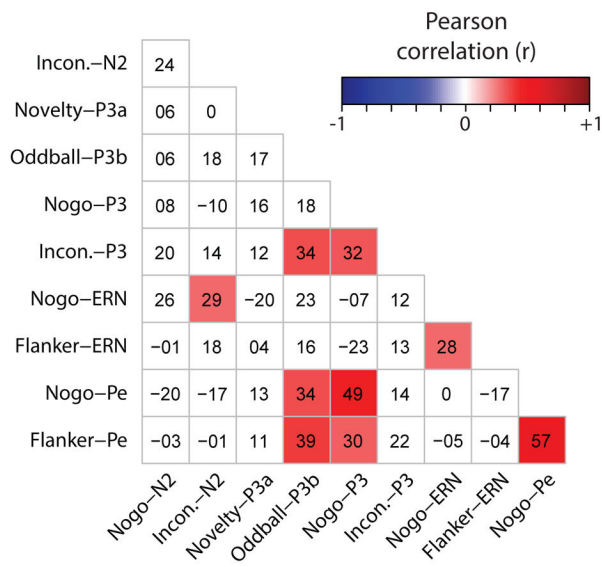


Figure 4.

A) Cross-measure correlation



B) Within-pair/cross-measure correlation

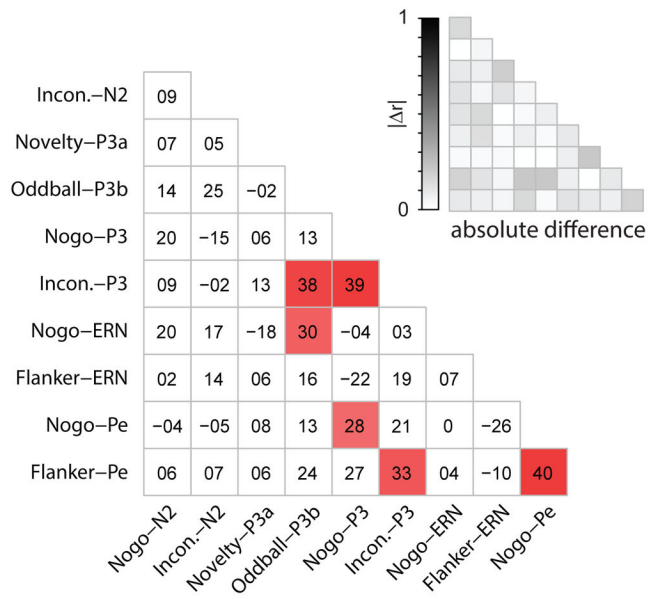


Figure 5.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 1**

Factor loadings for each of the three factors from principal components analysis.

	Factor loadings		
	1 (P3-F)	2 (N2-F)	3 (RN-F)
<i>Oddball/Novelty</i>			
Novelty-P3a	<b>.52</b>	.40	-.25
Oddball-P3b	<b>.70</b>	.29	.05
<i>Go/Nogo</i>			
Go-N2	.11	<b>.83</b>	.23
Nogo-N2	.04	<b>.73</b>	.32
Go-P3	<b>.76</b>	.14	.26
Nogo-P3	<b>.84</b>	.15	.17
Go-CRN	.24	.15	<b>.78</b>
Nogo-ERN	-.03	.22	<b>.68</b>
Nogo-Pe	<b>.72</b>	-.06	.15
<i>Flanker</i>			
Flanker-N2	.07	<b>.84</b>	.07
Flanker-P3	<b>.85</b>	.09	.12
Flanker-CRN	.36	-.12	<b>.75</b>
Flanker-ERN	.09	.35	<b>.62</b>
Flanker-Pe	<b>.62</b>	-.29	.10

Note: Each individual ERP measure's highest loading among the three factors is presented in boldface. As is evident from the pattern of loadings, the first factor primarily explains variance in P3 and Pe measures (hereinafter referred to as P3-F), the second explains variance in N2 measures (N2-F), and the third explains variance in response-negativity measures (ERNs and CRN; RN-F).

Descriptive statistics for time 1 (T1) and time 2 (T2) (left) and age-related linear mixed-effects (LME) regression results (right) for task accuracy, reaction time, and ERP amplitude measures for the Oddball/Novelty, Go/Nogo, and Flanker tasks.

**Table 2**

	Group means				Effect of age LME regression			
	T1		T2		B <sub>2</sub>		SE	
	Mean	SD	Mean	SD	B <sub>2</sub>	SE	SE	P
<b>Accuracy (%)</b>								
Oddball/Novelty	94.19	9.15	96.63	5.96	2.20	.59		<.001 *
Go/Nogo	91.80	6.20	94.93	4.50	2.79	.36		<.001 *
Flanker	93.20	7.79	96.30	5.19	2.51	.52		<.001 *
<b>Reaction time (milliseconds)</b>								
Oddball/Novelty	902.78	128.55	866.75	125.66	-27.61	9.37		.004 *
Go/Nogo	417.71	80.11	448.99	79.66	24.80	5.50		<.001 *
Flanker	526.97	75.53	540.50	67.04	10.29	5.40		.060
<b>Amplitude (microvolts)</b>								
<i>Oddball/Novelty</i>								
Novelty-P3a	5.75	8.74	4.87	7.76	-.58	.50		.254
Oddball-P3b	18.88	7.86	16.10	6.90	-2.98	.47		<.001 *
<i>Go/Nogo</i>								
Go-N2	-.53	4.21	-.62	3.63	-.16	.24		.509
Nogo-N2	-1.08	5.68	-1.36	5.66	-.15	.28		.608
Go-P3	8.12	3.51	6.18	3.23	-1.74	.23		<.001 *
Nogo-P3	14.37	5.70	11.68	5.37	-2.25	.39		<.001 *
Go-CRN	1.13	4.13	.12	3.97	-.77	.32		.018
Nogo-ERN	-5.51	6.57	-6.59	7.40	-1.21	.56		.032
Nogo-Pe	12.16	8.36	10.88	10.61	-1.34	.68		.052
<i>Flanker</i>								
Congruent-N2	-1.85	4.04	-1.33	3.79	.23	.27		.400
Incongruent-N2	-1.96	4.09	-1.15	3.83	.43	.27		.124

	Group means				Effect of age LME regression			
	T1	T2	T1	T2	B <sub>2</sub>	SE	p	
	Mean	SD	Mean	SD				
Congruent-P3	13.30	4.69	10.55	4.68	-2.41	.32	<.001*	
Incongruent-P3	15.14	5.35	12.74	5.35	-2.19	.35	<.001*	
Flanker-CRN	2.78	4.77	1.69	4.88	-.74	.36	.042	
Flanker-ERN	-.88	5.76	-2.21	6.62	-1.27	.50	.013	
Flanker-Pe	17.40	9.53	14.61	11.51	-1.91	.87	.030	
<b>ERP factors (standard deviations)</b>								
P3-F	.24	.99	-.25	.96	-.46	.06	<.001*	
N2-F	-.05	.95	.01	1.00	.07	.05	.192	
RN-F	.14	.94	-.06	1.01	-.16	.07	.019	

Note: The three bottom-most rows display values for factors extracted using principal components analysis reflecting primary dimensions of variance in ERP amplitude scores (see Methods). Of the forty-eight families that participated across the two sessions (mean age at T1 = 15.5 years [SD = .9]; T2 = 16.4 years [SD = .9]), available ERP scores (see text) for the different tasks and accuracies differed slightly: Go/Nogo incorrect at T1 = 92; Go/Nogo correct at T1 = 92; Go/Nogo incorrect at T2 = 95; Go/Nogo correct at T2 = 96; Flanker incorrect at T1 = 91; Flanker correct at T1 = 94; Flanker incorrect at T2 = 93; Flanker correct at T2 = 96. Coefficients (B<sub>2</sub>) from LME models depict the average rate of change associated with a one-year increase in age, accompanied by standard errors (SEs) and significance-values (p).

\* Indicates the p-value falls below the .01 cutoff adopted for this study.

**Table 3**

Comparison of using ERP scores extracted from raw voltage waveforms versus difference waveforms.

	Developmental stability ( <i>r</i> )			Twin similarity (ICC)		
	Contrast: Difference wave	Raw voltage measure	Diff. wave measure	Difference minus raw voltage	Raw voltage measure	Diff. wave measure
<i>Oddball/Novelty task</i>						
P3a	P3a-Standard	.81	.78	-.03	.69	.64
P3b	P3b-Standard	.77	.72	-.05	.63	.62
<i>Go/Nogo task</i>						
Nogo-N2	Nogo-Go	<b>.88</b>	<b>.74</b>	<b>-.14</b>	.64	.53
Nogo-P3	Nogo-Go	<b>.75</b>	<b>.64</b>	<b>-.11</b>	.54	.59
ERN	Error-Correct	<b>.55</b>	<b>.34</b>	<b>-.21</b>	.63	.51
Pe	Error-Correct	.70	.63	-.07	.44	.42
<i>Flanker task</i>						
Incon.-N2	Incongruent-Congruent	<b>.75</b>	<b>.26</b>	<b>-.49</b>	<b>.82</b>	<b>.38</b>
Incon.-P3	Incongruent-Congruent	<b>.78</b>	<b>.51</b>	<b>-.27</b>	<b>.77</b>	<b>.42</b>
ERN	Error-Correct	.45	.41	-.04	.54	.49
Pe	Error-Correct	.56	.54	-.02	.51	.56

Developmental stability (Pearson correlation, or *r*) for the 1 year interval spanning time 1 and time 2 and twin similarity (intraclass correlation, or ICC) among twin pairs is presented using either scores extracted from raw voltage ERPs (derived directly from scalp recordings) or difference waveform ERPs (derived from contrasts between two raw voltage ERP waveforms). The contrasts between task manipulations is presented in the second column labeled "Contrast: difference wave." All correlations are significant at  $p < .01$ . Numerical change in *r* or ICC is also presented ("Difference minus raw voltage") whereby negative values suggest nominally weaker stability or twin similarity, respectively. Instances where *r* or ICC for a given difference wave measure falls below the 95% confidence interval of its respective raw voltage measure in Figure 2 are highlighted in bold and indicate a significant reduction in stability or twin similarity.