Brain-behavior relationships in externalizing: P3 amplitude reduction reflects deficient inhibitory control

Grace M. Brennan*, Arielle R. Baskin-Sommers

Department of Psychology, Yale University, United States

**ABSTRACT**

The use of endophenotypes to classify individuals at risk for or suffering from psychopathology has been criticized for lacking specificity and predictive utility. This issue is apparent in research on externalizing, a heritable predisposition to disinhibitory psychopathology and personality traits. Numerous studies have shown that P3 amplitude reduction (P3AR) reliably reflects externalizing, implicating P3AR as a candidate endophenotype for externalizing psychopathology. However, this endophenotype has not been connected directly to a key deficit in executive function (e.g., inhibitory control) commonly related to externalizing. Using a modified oddball task in a sample (N = 74) of at-risk adolescents and young adults, we examined the associations among externalizing, P3AR, and inhibitory control. We also examined the associations of P3AR and inhibitory control with frequency of real-world disinhibited behavior. Results indicated that externalizing related to P3AR, which in turn related to deficient inhibitory control. Additionally, there were both unique and interactive associations of P3 amplitude and inhibitory control with indicators of real-world behavior. These findings provide the first direct evidence that P3AR reflects deficits in inhibitory control, thus linking this externalizing-related endophenotype to a specific cognitive process. Moreover, the results highlight the value of considering psychobiological measures alongside behavioral measures for indexing risk for externalizing behavior and psychopathology.

1. Introduction

Externalizing is a latent construct that represents a heritable predisposition to a broad spectrum of psychopathology (e.g., attention-deficit/hyperactivity disorder, conduct disorder, substance use disorders, antisocial personality disorder [1]) and personality traits (e.g., impulsivity, negative emotionality, low constraint [2]). Across these disorders and traits, a hallmark of externalizing is the pervasiveness of disinhibited behavior (e.g., rule-breaking, aggression, substance abuse). There is evidence that these disinhibited behaviors are underpinned by impairments in cognitive processes, specifically executive functions [3–5]. Notably, individuals with high levels of externalizing perform poorly on tasks that involve inhibitory control [6–15], a core executive function defined as the ability to suppress a prepotent (i.e., habitual, automatic) response in favor of a less automatic, task- or goal-relevant response [16]. Moreover, these externalizing-related inhibitory control deficits seem to emerge most readily in the context of salient stimuli [17–27].

A psychobiological marker purported to reflect the cognitive deficit in externalizing is reduced amplitude of the P3 event-related potential (ERP) [28]. In general, P3 is theorized to represent salience detection, stimulus evaluation, and updating of working memory [29]. Decades of research indicate that externalizing is negatively associated with P3 amplitude in the oddball paradigm, which involves responding to target stimuli that occur infrequently and unpredictably within a series of frequent stimuli. In this paradigm, participants are instructed to respond with a button press to infrequent stimuli but not to frequent stimuli. The P3 amplitude reduction (P3AR) to infrequent (“oddball”) stimuli is evident not only in individuals with externalizing disorders and traits [30], but also in individuals defined as at-risk on the basis of familial relationships [31,32]. Furthermore, the relation between P3 and proneness to externalizing psychopathology is mediated by heritable influences [33]. Together, these findings suggest that P3AR is an indicator of a common genetic vulnerability to disinhibitory pathologies and traits (i.e., it is an index of externalizing). Accordingly, P3AR has been widely acknowledged as a promising candidate endophenotype of externalizing psychopathology [34,35].

Despite the acceptance of P3AR as an endophenotype of externalizing, the precise processes reflected by P3AR remain unclear [36]. Although some researchers have attempted to draw a connection between executive dysfunction and P3AR, they have done so in an indirect manner. For example, Roca et al. [37] assessed P3 amplitude and

---

* Corresponding author at: Yale University, P.O. Box 208205 New Haven, CT 06520, United States.
E-mail address: grace.m.brennan@yale.edu (G.M. Brennan).

http://dx.doi.org/10.1016/j.bbr.2017.09.045
Received 15 July 2017; Received in revised form 19 September 2017; Accepted 27 September 2017
Available online 28 September 2017
0166-4328/ © 2017 Elsevier B.V. All rights reserved.
latency in children with ADHD and reported that P3 was related to deficits in father-reported executive function. Similarly, Kim et al. [38] found that P3AR in adolescents with conduct disorder was related to weaker cognitive control, assessed via Stroop task performance. Although relating separate measures (e.g., oddball task P3 with Stroop task performance) is a starting point for identifying the processes reflected by a psychobiological measure, a more direct examination involves concurrent measurement of P3 and behavioral performance, under the same conditions in the same task environment. This point is particularly germane to the study of externalizing, given evidence that externalizing-related executive function deficits are moderated by experimental context [39].

Another reason that the processes reflected by P3AR remain unclear is that studies using the traditional oddball paradigm (the paradigm most commonly used to assess P3AR in externalizing) demonstrate a lack of neural and behavioral concordance. Specifically, individuals with high levels of externalizing reliably exhibit P3AR on these tasks but do not show performance deficits (e.g., [30]). In the absence of observable differences in behavior, it is impossible to identify the precise cognitive processes associated with P3AR. Thus, a disconnect exists between the psychobiological research on P3AR on the one hand and the theoretical literature delineating the cognitive dysfunctions associated with externalizing on the other. The lack of concordance is likely due to the fact that the traditional oddball task is not designed to reveal externalizing-related performance differences related to inhibitory control. As noted above, in traditional oddball tasks, participants are instructed to respond with a button press to infrequent (and salient) stimuli and to not respond to frequent stimuli. Thus, the task requires participants to respond to salient information, thereby failing to impose demands on executive functions (e.g., inhibitory control). Although the task permits the assessment of psychophysiological responses to salient stimuli, it does not permit the evaluation of executive functions in the context of salience. Even so, researchers often interpret the externalizing-related P3AR as representing some etiologically relevant form of executive dysfunction (e.g., [40]). However, these interpretations are speculative and rely on reverse inference rather than direct empirical hypothesis testing. The resulting disconnect between the psychobiological and theoretical literature is problematic because it represents the gap between description and understanding of brain-behavior relationships; knowledge of either biological or behavioral processes alone is not sufficient for understanding the underlying processes of how the brain generates behavior [41]. Considering psychobiological and behavioral data jointly can thus enhance our understanding of the processes involved in complex human behavior. Again, this point is especially important for externalizing, given that the impairments associated with externalizing psychopathology are reflected in harmful behaviors such as reactive aggression, excessive use of substances, and physically dangerous risk-taking.

The primary aim of the present study was to determine whether P3AR is associated with behavioral performance when a demand on inhibitory control is imposed. Whereas traditional oddball paradigms required participants to respond to salient, infrequent stimuli, the present study’s modified oddball paradigm required participants to inhibit a response to salient, infrequent stimuli, thus placing a demand on inhibitory control. As a secondary aim, the present study explored whether task-based inhibitory control adds predictive value above and beyond P3AR in relation to the frequency of real-world disinhibited behavior. We used a sample of adolescents and young adults in order to assess P3AR at a developmental period before P3 amplitude differences have begun to diminish (e.g., [42]) and before prolonged, heavy substance use has altered the neurobiological functioning of individuals with externalizing. Moreover, the decision to include both adolescents and young adults in our sample was based on previous research indicating that the externalizing-related P3AR is evident not only in adolescents [30,32], but also in young adults [43]. We hypothesized that latent trait externalizing would predict P3AR on infrequent trials, which in turn would predict diminished accuracy on infrequent trials (i.e., deficient inhibitory control). Furthermore, we hypothesized that task-based inhibitory control would contribute to the prediction of real-world disinhibited behavior above and beyond task-related neural response (P3 amplitude), and that models incorporating both neural response and task-based behavior would be more informative than models considering neural response alone. Demonstrating that P3AR is connected to specific impairments in behavioral performance would allow assessment tools to be refined to reflect a more mechanism-based conceptualization of risk for externalizing psychopathology. This richer conceptualization, developed through the consideration of biological measures in the context of behavior [41], would enhance the predictive validity of these tools and enable more accurate prediction of risk for developing externalizing psychopathology. Furthermore, demonstrating that task-based inhibitory control predicts real-world disinhibited behavior above and beyond P3AR would underscore the value of considering behavioral measures in conjunction with psychophysiology.

2. Method

2.1. Participants

Participants were recruited from the New Haven community. The New Haven area is a high-crime region. Nationally, New Haven ranks in the 94th percentile for crime; on average, 344 crimes are committed per square mile, compared to the national median of 32.9 (Note: Data accessed from http://www.neighborhoodscout.com/ct/new-haven/crime/on/03/27/2017/). The rate of violent crime is 9.12 (per 1000 residents), compared to a statewide rate of 2.18 and a national median of 3.8. Additionally, the vast majority of participants (92.13%) endorsed a positive family history of substance use disorders. Thus, on the basis of being recruited from urban, high-crime regions [44,45] and having a family history of substance use disorders [46], the sample is at risk for externalizing behavior and psychopathology.

A prescreen interview was completed to exclude individuals who had a history of schizophrenia, bipolar disorder, or psychosis, not otherwise specified; a family history of psychosis; or a history of medical problems (e.g., uncorrectable auditory or visual deficits; head injury with loss of consciousness greater than 30 min) that may have impacted their comprehension of the materials or performance on the task. In the first session, participants provided written informed consent if 18 years of age or older, and assent/parental consent if under 18 years of age, in line with the procedures set forth by the Yale University Human Investigation Committee. They then completed the Shipley Institute of Living Scale [47] (see Materials and Measures subsection below for details), which provides an estimate of IQ, and completed several self-report measures of personality and behavior. Participants with an estimated IQ of 70 or above were eligible to continue. During the second session, participants completed the experimental task. Participants were paid $30 per session.

Participants were 59 males and 30 females between the ages of 14 and 24 ($M = 19.65, SD = 2.93$). In terms of race, the majority of participants were African American (68.5%), while the remaining participants self-identified as mixed racial identity (18%), White (11.2%), or other (2.2%). 13.5% of participants self-identified as Hispanic.

2.2. Materials and measures

2.2.1. Alcohol use disorders identification test (AUDIT; [48])

The AUDIT is a 10-item self-report questionnaire used to identify hazardous and harmful patterns of alcohol consumption. Respondents rate each item on a 5-point Likert scale. The total score is acquired by summing ratings for each item. Total scores can range from 0 to 40, with higher scores reflecting higher levels of alcohol-related problems. Psychometrically, this instrument has demonstrated high internal
consistency, test-retest reliability, convergent validity, sensitivity, and specificity [49]. It has been validated for use in adolescents [50, 107, 51]. For this sample, internal consistency (i.e., reliability) was in the acceptable range (Cronbach’s α = 0.65); a lower α compared to our other self-report measures is to be expected given the small number of items in the measure. Furthermore, this α value is not outside the range of published and expected reliabilities for this measure [52].

2.2.2. Drug abuse screening test for adolescents (DAST-A; [53])

The DAST-A was adapted from the Drug Abuse Screening Test (DAST; [54]) for use as a screening tool in adolescents. It is a 27-item self-report questionnaire used to assess problems related to drug use. Respondents rate each item as either “Yes” or “No.” The total score is computed by summing all items that are endorsed in the direction of increased drug-related problems. Total scores can range from 0 to 27, with higher scores indicating higher levels of drug-related problems. For this sample, good internal consistency (i.e., reliability) was demonstrated (Cronbach’s α = 0.82).

2.2.3. Sensation Seeking Survey Form V seeking scale V (SSS-V; [55])

The SSS-V is a 40-item self-report measure that assesses a person’s overall propensity for seeking out novel and stimulating experiences. The measure contains four subscales, each containing ten items, which assess four specific facets of sensation seeking: Thrill and adventure seeking (a tendency to engage in physically dangerous activities), Experience seeking (desire for varied sensory and life experiences), Disinhibition (tendencies toward social, sexual, and substance-related disinhibition), and Boredom susceptibility (difficulty tolerating monotony). For each item, respondents select the one option (out of two) that better describes their preferences and feelings. We selected the Disinhibition subscale of the SSS-V as a measure of low constraint, which comes from the Multidimensional Personality Questionnaire (MPQ [56]) and captures a tendency to prefer spontaneous, behavioral, and unconventional activities. Scores on the Disinhibition subscale range from 0 to 10, with higher scores reflecting higher levels of disinhibition. For this sample, acceptable internal consistency (i.e., reliability) was demonstrated (Cronbach’s α = 0.59). Again, a lower α compared to our other self-report measures is to be expected given the small number of items in the subscale.

2.2.4. Risky, impulsive, self-destructive questionnaire (RISQ; [57])

The RISQ is a 38-item self-report questionnaire that measures risky, impulsive, and self-destructive behaviors in eight domains: aggression, self-harm, gambling, reckless behavior, impulsive eating, risky sex, drug use, and alcohol use. For each behavior, respondents note the number of times they have engaged in the behavior in their lifetime, how many times in the past month, and how old they were when they first started engaging in the behavior. Additionally, respondents indicate if there were any consequences (e.g., legal, social, financial) as a result of their behavior. Finally, for each behavior respondents indicate how strongly they agree with statements that assess their motivation (distress relief or pleasure seeking) for engaging in the behavior. Psychometrically, this instrument has demonstrated high internal consistency and construct validity [57].

2.2.5. Beck depression inventory-II (BDI-II; [58])

The BDI-II is a 21-item self-report questionnaire assessing severity of depressive symptoms in the past 2 weeks. Respondents are asked to rate individual symptoms (e.g., sadness, pessimism, irritability, perception of failure) on a 4-point Likert scale from 0 (does not identify with symptom) to 4 (strongly identifies with symptom). Higher scores correspond with greater severity of depressive symptoms, and standard score ranges have been determined for minimal depression (0–9), mild depression (10–18), moderate depression (19–29), and severe depression (30–63). For this sample, excellent internal consistency (i.e., reliability) was demonstrated (Cronbach’s α = 0.93).

2.2.6. State-Trait anxiety inventory (STAI; [59])

The STAI is a 40-item self-report measure that assesses anxiety in terms of state- and trait-based levels of stress. For the present study, only the trait-anxiety subscale was examined. The trait-anxiety subscale measures anxiety in terms of a persistent lifetime personality trait (e.g., “I take disappointments so keenly that I can’t put them out of my mind”). These 20 questions are rated on a 4-point Likert scale from 0 (almost never) to 3 (almost always). Higher scores on this subscale indicate higher levels of trait-based anxiety. For this sample, acceptable internal consistency (i.e., reliability) was demonstrated (Cronbach’s α = 0.64).

2.2.7. Shipley institute of living scale [47]

The Shipley is a measure of intelligence that consists of two subtests: vocabulary, a 40-item subtest in which participants choose a word (out of four options) that is synonymous with the word provided; and pattern matching, a 20-item subtest in which participants complete verbal and numerical patterns by writing in correct answers. Examiners convert raw scores on each subtest and then the total raw score to age-corrected T-scores. The total age-corrected T-score can then be used to estimate a participant’s WAIS-R Full-scale IQ score, which has been shown to be an accurate means of predicting IQ [60].

2.2.8. Modified oddball task

The modified oddball task was an adaptation of the rotated-heads visual oddball paradigm [31]. During the task, participants were presented with frequent and infrequent stimuli. Participants were instructed to respond with a button press each time a circle (frequent stimulus) appeared on the screen and to withhold a response each time a head (infrequent stimulus) appeared on the screen (see Fig. 1). Participants were further instructed to respond as quickly and accurately as possible. The task consisted of 240 trials; frequent stimuli (circles) appeared on two-thirds of the trials, and infrequent stimuli (heads) appeared on one-third of the trials, consistent with stimulus frequencies used in previous studies (e.g., [32]).

Stimulus presentation and response collection were controlled using the Psychtoolbox extension [61–63] as implemented in Matlab.

Frequent stimuli:

Infrequent stimuli:

Fig. 1. Schematic depicting the categories of stimuli displayed to participants. Participants were instructed to respond with a button press when a circle (top) appeared, which occurred on 160 trials (two-thirds of trials). Participants were instructed to withhold a button press when a head (bottom) appeared, which occurred on 80 trials (one-third of trials; each of the four orientations appeared in 20 trials).
G.M. Brennan, A.R. Baaske-Sommers

the EEGLAB toolbox[65] in MATLAB. As an initial step, 60-Hz noise electrodes. EEG data were digitally components. EEG data were re-referenced to an average reference of all times (fi indicated a signifi midline parietal/occipital sites had the lowest 60-Hz noise. The results linear model (GLM) analysis was conducted to identify which of three amplitude was most cleanly measurable. A repeated-measures general

2.3. Psychophysiological recording and analysis

EEG was recorded throughout the experiment from 128 Ag/AgCl electrodes embedded within a HydroCel Geodesic sensor net, using NetStation v.4.2 software (Electrical Geodesics, Incorporated [EGI]) and EGI high-impedance amplifiers, sampled at 1000 Hz (1 Hz high-pass, 100 Hz low-pass). All electrodes were referenced to Cz for recording. Electrooculogram (EOG) was recorded above and below the key variables for the overall sample. Excluded participants did not differ significantly from included participants in terms of any of the following characteristics: sex, age, IQ, AUDIT score, DAST-A score, or SSS Disinhibition score (all p values ≥ 0.078). The final analyzed sample of 74 participants consisted of 52 males (70.27%) and 22 females (29.73%). Sample characteristics, descriptive statistics, and correlations among key variables for the final sample are presented in Table 1.

| Variable | Mean | SD | Range | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----------|------|----|-------|---|---|---|---|---|---|---|---|---|----|----|----|----|
| 1) Age   | 19.81| 2.93| 14-24 | - | - | - | - | - | - | - | - | - | - | - |
| 2) IQ    | 97.78| 9.32| 72-116| -0.08 | - | - | - | - | - | - | - | - | - | - |
| 3) AUDIT | 1.97 | 2.53| 0-11  | 0.34**| 0.37**| - | - | - | - | - | - | - | - | - |
| 4) DAST-A| 1.74 | 2.62| 0-10  | 0.15 | 0.17 | 0.30**| - | - | - | - | - | - | - | - |
| 5) SSS Disinhibition | 4.55 | 2.22| 0-9   | 0.19 | 0.22* | 0.47**| 0.40**| - | - | - | - | - | - | - |
| 6) SSS Total | 16.68| 5.49| 4-31  | 0.11 | 0.37**| 0.44**| 0.34**| 0.69**| - | - | - | - | - | - |
| 7) RISQ Drug | 28.19| 130.17| 0-1100 | 0.11 | 0.15 | 0.30**| 0.30**| 0.14 | 0.16| - | - | - | - | - |
| 8) RISQ Aggression | 1.41| 6.30| 0-50  | -0.31**| 0.03 | -0.12 | 0.08 | -0.08 | -0.18 | -0.03| - | - | - | - |
| 9) RISQ Reckless | 5.20| 16.95| 0-105 | 0.10 | 0.10 | 0.37**| 0.02 | 0.17 | 0.22*| 0.05 | 0.01| - | - | - |
| 10) RISQ Total | 37.59| 131.16| 0-1100 | 0.11 | 0.18 | 0.39**| 0.26*| 0.15 | 0.18 | 0.95**| 0.06 | 0.33**| - | - |
| 11) “Go” trial accuracy | 0.95 | 0.05| 0.78-1.00 | 0.29**| 0.18**| 0.04 | 0.13 | 0.12 | -0.02 | 0.13 | 0.07 | 0.11 | 0.16| - |
| 12) “No-go” trial accuracy | 0.71 | 0.20| 0.13-0.97 | 0.38**| 0.20*| 0.13 | -0.01 | 0.07 | -0.06 | -0.16 | -0.10 | 0.01 | -0.13 | 0.32**| - |
| 13) “Go” trial reaction time | 0.35 | 0.07| 0.14-0.47 | 0.29**| 0.11 | 0.14 | -0.01 | 0.05 | 0.00 | -0.22**| -0.01 | -0.12 | -0.23*| 0.17 | 0.81**| - |

Note. IQ = WAIS IQ estimate from the Shipley Institute of Living Scale, AUDIT = Alcohol Use Disorders Identification Test total score, DAST-A = Drug Abuse Screening Test for Adolescents total score, SSS Disinhibition = Disinhibition subscale score from the Sensation Seeking Scale, SSS Total = Sensation Seeking Scale total score, RISQ Drug = number of times participant engaged in drug-related behavior in the past month, RISQ Aggression = number of times participant engaged in aggression in the past month, RISQ Reckless = number of times participant engaged in reckless behavior in the past month, RISQ Total = number of times participant engaged in a disinhibited behavior in the past month, summed across the 8 behavior domains included in the Risky, Impulsive, Self-Destructive Questionnaire. * p < 0.1, ** p < 0.05.

Data analysis occurred in four stages. First, in order to represent latent trait externalizing, we used structural equation modeling (SEM) in Stata (StataCorp) to derive a measurement model based on that of Krueger et al. [1]. SEM is a form of multivariate analysis that evaluates the overall fit of a proposed model of the associations among latent (i.e., unobserved) and observed variables [68]. This statistical technique was chosen as well-suited to the goals of the present study due to its capacity to estimate overall model fit and account for sources of error [69]. The observed variables used to estimate the latent externalizing variable were alcohol-related problems (AUDIT total score), drug-related problems (DAST-A total score), and low constraint (Disinhibition subscale of the SSS-V). Second, a series of repeated measures GLM analyses was conducted to examine whether there were differences in P3 amplitude and behavioral performance based on trial type. In the first model, P3 amplitude
served as the continuous dependent measure, and trial type (frequent/"go" versus infrequent/"no-go") was a categorical within-subjects repeated measure. Externalizing score was entered as a continuous predictor. In the second model, accuracy served as the continuous dependent measure, and trial type was a categorical within-subjects repeated measure. Externalizing score was entered as a continuous predictor. To protect against violations of the assumption of sphericity, Huynh-Feldt corrected $p$ values are reported.

Third, the associations among externalizing, P3 amplitude, and behavioral performance on the modified oddball task were assessed using a structural model, which was built to test the set of hypotheses that externalizing would predict lower P3 amplitude, and lower P3 amplitude would in turn predict lower accuracy on infrequent "no-go" trials (i.e., weaker inhibitory control). Age and IQ were entered as covariates in the model. The choice to use a full structural model to test the associations among externalizing, P3 amplitude, and behavioral performance was made for several reasons. Through a confirmatory approach [70], which involves constructing a theory-driven model, SEM has the capacity to test a series of hypothesized associations within a single coherent model, which is more advantageous than calculating associations in a piecemeal fashion (e.g., using a series of separate regression analyses). Additionally, SEM can provide an overall measure of model fit in the context of multiple hypothesized associations. Tests of goodness of fit included the relative chi-square index (the ratio of the chi-square statistic to the degrees of freedom), comparative fit index (CFI) [71], and the Root Mean Square Error of Approximation (RMSEA) [72] and its corresponding $p$ of close fit (PCLOSE). Traditionally, relative chi-square index values $< 5$ [73], CFI $> 0.90$ [74], RMSEA $\leq 0.05$ [75], and PCLOSE $\geq 0.05$ are considered signs of good fit. Finally, SEM is capable of testing mediational (i.e., path) hypotheses, and does so using a flexible and comprehensive framework [76]. The advantage of using SEM over other methods (e.g., bootstrapping) to test mediational hypotheses lies in the fact that SEM is capable of modeling residual error (i.e., error in the prediction of one variable to another); traditional multivariate procedures (e.g., regression) cannot model and account for residual error. Thus, because we sought to test hypothesized associations, assess overall model fit, and account for sources of error, a full SEM method was implemented.

Lastly, the associations of task-based neural response and task-based performance with frequency of self-reported past-month disinhibited behavior were assessed using a series of negative binomial regression models. We tested both main effects and interactions of neural and behavioral measures in predicting various forms of real-world disinhibited behavior. We included behavioral categories of the RISQ when at least 10% of participants reported engaging in the behavior at least once in the past month. These categories were drug use (40.54% endorsed at least once), aggression (13.51%), gambling (17.57%), risky sex (16.22%), and reckless behaviors (39.19%). Only significant results are presented for these category-related analyses. We then used Akaike Information Criterion (AIC) values for models with the neural predictor (P3AR) alone versus neural and behavioral (task-based “no-go” accuracy) predictors to assess whether inclusion of the task-based behavioral measure enhanced the prediction of real-world disinhibited behavior. The AIC is a measure of the relative quality of statistical models, offering a relative estimate of the information lost when a given model is used to represent the process that generates the data [77]. Thus, the use of AIC provides a basis for model selection, with lower values indicating better model fit.

3. Results

3.1. Task-related effects on behavioral performance

A total of 240 stimuli were presented: 80 infrequent/"no-go" (heads) and 160 frequent/"go" (circles). Accuracy was significantly higher on "go" versus "no-go" trials, $F(1, 72) = 114.00, p < 0.001$, $\eta^2 = 0.61$ (see Table 1 for descriptive statistics and zero-order correlations among task variables).

3.2. Task-related effects on P3 amplitude

There was a significant main effect of trial type on P3 amplitude, such that infrequent/"no-go" trials elicited a larger P3 than frequent/"go" trials, $F(1, 72) = 30.50, p < 0.001$.

$\eta^2 = 0.30$. The mean P3 amplitude on infrequent/"no-go" trials was 4.03 $\mu V$ ($SD = 2.64$), and the mean P3 amplitude on frequent/"go" trials was 2.92 $\mu V$ ($SD = 2.19$) (see Fig. 2).

3.3. Externalizing, P3, and inhibitory control

The measurement model used to derive the latent externalizing variable indicated good fit ($\chi^2 = 11.02, CFI = 0.99, RMSEA = 0.00, PCLOSE = 1.00$), and all observed variables loaded significantly onto the latent externalizing factor (all $p$ values $< 0.001$; see Fig. 3).

The overall structural model (including both measurement model and path model) indicated good fit ($\chi^2 = 11.02, \frac{\chi^2}{df} = 1.00, CFI = 1.00, RMSEA = 0.005, PCLOSE = 0.60$), and all paths were significant (all $p$ values $< 0.05$; see Fig. 4). The first path, from the latent externalizing variable to P3 amplitude on "no-go" trials, was significant, $\beta = -0.28, p = 0.030, 95\% CI [-0.54, -0.03]$. Consistent

![Fig. 2. Average ERP waveform at POz for individuals low and high on externalizing latent variable scores. Though the primary analyses were conducted using continuous externalizing scores, “low EXT” (individuals who scored below the median on the externalizing latent variable) and “high EXT” (individuals who scored above the median on the externalizing latent variable) groups were used here solely for depiction. Blue lines represent the low EXT group, and red lines represent the high EXT group. Solid lines represent frequent/“go” trials, and dashed lines represent infrequent/“no-go” trials. The gray box indicates the timeframe used to derive the P3 amplitude measure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image)

![Fig. 3. Measurement model for the externalizing latent variable. Path coefficients are standardized beta weights and are all significant at $p < 0.001$. EXT = externalizing latent variable, DRUG = drug-related problems (DAST-A total score), ALC = alcohol-related problems (AUDIT total score), LCON = low constraint (SSS Disinhibition subscale score).](image)
disinhibited behavior (RISQ Past Month count) was entered as the dependent variable, P3 amplitude (on infrequent/“no-go” trials) and task “no-go” accuracy were entered as standardized continuous predictors, and age was entered as a standardized continuous covariate (since both risky behaviors and inhibitory control were positively correlated with age). Outliers in the behavior count variables (3 SDs or more above the mean) were winsorized. For each behavior category, there were no more than two outliers.

The model using a total count of disinhibited behaviors (across all eight behavior categories) as a dependent variable demonstrated good fit, \( \chi^2/df = 2.77, p < 0.001 \). Both P3, OR = 0.65, \( p = 0.008, 95\% CI [0.48, 0.89] \), and “no-go” accuracy, OR = 0.63, \( p = 0.005, 95\% CI [0.46, 0.87] \), were significantly related to frequency of disinhibited behavior. Specifically, participants with lower P3 amplitude engaged in disinhibited behaviors more frequently than participants with higher P3 amplitude. Additionally, participants with poorer inhibitory control engaged in disinhibited behaviors more frequently than participants with stronger inhibitory control. The interaction between P3 amplitude and “no-go” accuracy was not significant, \( p = 0.652 \).

In the model using past-month drug behaviors as a dependent variable (model fit: \( \chi^2/df = 3.83, p < 0.001 \)), both P3, OR = 0.51, \( p < 0.001, 95\% CI [0.37, 0.71] \), and “no-go” accuracy, OR = 0.38, \( p < 0.001, 95\% CI [0.26, 0.55] \), contributed significantly. Specifically, participants with lower P3 amplitude engaged in drug-related behaviors more frequently than participants with higher P3 amplitude. Additionally, participants with poorer inhibitory control engaged in drug-related behaviors more frequently than participants with stronger inhibitory control. The interaction between P3 amplitude and “no-go” accuracy was not significant, \( p = 0.260 \).

In the model using aggression as a dependent variable (model fit: \( \chi^2/df = 2.61, p < 0.001 \)), neither P3 nor inhibitory control was significantly related. However, the interaction between P3 amplitude and “no-go” accuracy was significant, \( OR = 2.41, p = 0.001, 95\% CI [1.43, 4.07] \), suggesting that, with a 1 standard deviation decrease in P3, the effect of inhibitory control decreased by 141%. At one standard deviation below the mean on P3 amplitude, worse inhibitory control was related to higher frequency of aggression; however, at one standard deviation above the mean on P3 amplitude, better inhibitory control was related to higher frequency of aggression.

In the model using reckless behaviors as a dependent variable (model fit: \( \chi^2/df = 2.78, p = 0.002 \)), neither P3 amplitude, \( p = 0.082 \), nor inhibitory control contributed significantly, \( p = 0.689 \). However, the interaction between P3 amplitude and “no-go” accuracy was significant, \( OR = 0.55, p = 0.001, 95\% CI [0.38, 0.79] \), suggesting that, with a 1 standard deviation decrease in P3, the effect of inhibitory control increased by 45%. At one standard deviation below the mean on P3 amplitude, better inhibitory control was related to higher frequency of reckless behavior; however, at one standard deviation above the mean on P3 amplitude, there was no difference in reckless behavior based on inhibitory control.

To assess whether models incorporating both task-based neural response and task-based performance (neural/behavioral model) demonstrated superior fit compared to models incorporating task-based neural response alone (neural model), the AIC was derived and compared across both types of models for each RISQ category examined. These results should be considered in light of the fact that the AIC penalizes a higher number of parameters (i.e., it favors parsimony) (see Table 3). Generally, delta AIC < 2 suggests substantial evidence for model i (in this case the neural model), values between 3 and 7 indicate that model i hasn’t considerably less support, whereas values > 10 indicate that model i is very unlikely [81]. For all models, the combination of neural and behavioral data demonstrated superior fit compared to models with the neural measure alone.

3.4. Behavior and psychophysiology in the prediction of real-world disinhibited behavior

In a series of negative binomial regression models, past-month...
Table 3

<table>
<thead>
<tr>
<th>Response Variable</th>
<th>Neural Model AIC</th>
<th>Neural/Behavioral Model AIC</th>
<th>Delta AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RISQ Total</td>
<td>636.78</td>
<td>632.70</td>
<td>4.08</td>
</tr>
<tr>
<td>RISQ Drug</td>
<td>555.82</td>
<td>530.02</td>
<td>25.80</td>
</tr>
<tr>
<td>RISQ Agression</td>
<td>165.86</td>
<td>158.12</td>
<td>7.74</td>
</tr>
<tr>
<td>RISQ Reckless</td>
<td>355.07</td>
<td>346.86</td>
<td>8.21</td>
</tr>
</tbody>
</table>

Note. AIC = Akaike Information Criterion; Delta AIC = A measure to compare models (AIC_neural-AICbehavioral). RISQ Total = number of times participant engaged in a disinhibited behavior in the past month, summed across the 8 behavior domains included in the Risky, Impulsive, Self-Destructive Questionnaire; RISQ Drug = number of times participant engaged in a drug-related behavior in the past month; RISQ Agression = number of times participant engaged in aggressive behavior in the past month; RISQ Reckless = number of times participant engaged in reckless behavior in the past month.

3.4.1. Supplemental analysis for behavior and psychophysiology in the prediction of real-world disinhibited behavior

To test the specificity of these associations to externalizing, we ran an additional negative binomial regression with internalizing symptoms (BDI-II score) as the dependent measure. The model did not demonstrate good fit, $p = 0.377$, indicating that our neural and behavioral measures were not associated with internalizing psychopathology.

4. Discussion

4.1. Interpretation and significance of findings

The primary aim of the present study was to examine the relationships among externalizing, P3AR, and inhibitory control. In order to assess the direct association between a candidate psychobiological endophenotype, P3AR, and a theoretical mechanism of externalizing, executive dysfunction, the traditional oddball task was modified to require participants to inhibit rather than emit a response to infrequent, salient stimuli. It follows that participants needed to deploy inhibitory control to inhibit the prepotent response (button press) in the context of salient stimuli. Consistent with hypotheses and previous research [28,30], results indicated that externalizing predicted lower P3 amplitude. Moreover, P3AR predicted lower accuracy on “no-go” trials. This finding represents the first direct demonstration that externalizing-related P3AR reflects executive dysfunction, specifically inhibitory control in the context of salience. Furthermore, both P3 amplitude and inhibitory control uniquely and interactively contributed to predicting frequency of real-world disinhibited behavior.

The novel finding that P3AR directly reflects impaired inhibitory control in the context of salience represents a conceptual and practical advance in our understanding of externalizing. Conceptually, the present study integrates two lines of research that have separately documented psychophysiological and cognitive processes involved in externalizing. Illustrating this brain-behavior relationship brings clarity to the functional significance of P3AR for externalizing, which has long been unknown [79]. Considering the present results in the context of prior work that has found P3AR in the absence of performance differences, it is possible that externalizing involves cognitive abnormalities associated with processing salient stimuli, which lead to deleterious downstream consequences for the inhibition of prepotent responses. Since there are multiple cognitive components supporting inhibitory control (e.g., maintaining task rules, vigilance, sustained attention, motor inhibition [80]), it is possible that any or all of these components are disrupted in externalizing. However, since the inclusion of “go”-trial accuracy as a covariate in the SEM did not affect our results, it seems rather unlikely that maintaining task rules or sustained attention are responsible for the inhibitory control deficits seen in the present study, as these processes would have impaired task performance across the board. Instead, the results of the present study suggest an externalizing-related deficit in exercising inhibitory control in response to salient information.

The second research question addressed whether laboratory-based behavioral measures can serve as incrementally informative tools (i.e., above and beyond psychobiological measures) for understanding and potentially predicting engagement in disinhibited behaviors. For each category of real-world disinhibited behavior, the model incorporating both neural and behavioral measures demonstrated better fit than the model based on neural activity alone (see Table 3 above). However, the specific patterns of associations diverged across behaviors.

Consistent with previous research [82,10], both P3AR and inhibitory control were uniquely related to frequency of disinhibited behavior in general as well as drug use-related behavior. Moreover, this result adds to previous findings by indicating that both measures, when considered simultaneously, uniquely predict the frequency of these behaviors. Thus, although in the present study P3AR and inhibitory control were related and appear to overlap to some extent, each measure provides unique information in terms of associations with disinhibited behaviors.

Conversely, neither P3AR nor inhibitory control was independently associated with frequency of aggressive or reckless behavior. Rather, they interacted to predict the frequency of these types of behavior, demonstrating that the influence of each index is dependent on the level of the other. Consistent with previous research suggesting that deficient inhibitory control is not necessarily a key factor underlying aggression [83], the importance of inhibitory control for frequency of aggressive behavior decreased as P3 amplitude decreased. Although the combination of lower P3 amplitude and poorer inhibitory control was related to higher frequency of aggression, surprisingly individuals who engaged in aggression most frequently had higher P3 amplitude and better inhibitory control. This interaction may be explained by the fact that there are different forms of aggression (e.g., reactive versus proactive) that vary based on personality traits and exhibit dissociable patterns of P3 amplitude [84–86] and executive function capabilities [87]. For example, for a subtype of antisocial individuals (i.e., those with callous-unemotional traits), better executive function predicts higher levels of aggression [88]. An additional consideration is that we did not measure P3 or inhibitory control in a threat-related context, which may be crucial for detecting deficits related to reactive aggression [89]. Although the present study did not distinguish between subtypes of aggression or manipulate affective context, future research could investigate their associations with P3 amplitude and inhibitory control.

For reckless behavior, by contrast, the importance of inhibitory control increased as P3 amplitude decreased. Individuals who engaged in reckless behavior more frequently had lower P3 amplitude and better inhibitory control. This interaction may be related to the fact that reckless behavior tracks levels of sensation-seeking (see Table 1), which normatively peaks during adolescence [90] and is a trait distinct from impulsivity [91]. Thus, consistent with previous research [92], deficits in inhibitory control were not related to sensation-seeking behaviors in the present study. Taken together, these findings across behaviors support the idea that the integration of multiple levels of analysis (i.e., neural, behavioral) is useful when seeking to understand complex behaviors and pathologies [93]. Thus, it stands to reason that the predictive utility of P3AR as an externalizing endophenotype could be enhanced by simultaneously considering behavioral measures (e.g., inhibitory control) so as to examine unique and interactive effects of both types of measures.

Currently, P3AR alone cannot differentially predict risk for forms of psychopathology that are divergent in terms of etiology, clinical symptoms, and indicated treatments (e.g., internalizing versus externalizing [94,95]). This is detrimental because endophenotypes would be more valuable if they could be harnessed to predict an individual’s risk for developing a circumscribed range of disorders (and not just psychopathology in general [96,35]). The present study...
suggests that it is possible to enhance the specificity of P3AR, in terms of predicting risk for psychopathology, by considering it alongside behavioral measures, thus taking advantage of additional input than can enhance our ability to detect the presence of key cognitive processes (see Table 3). Thus, whereas P3AR alone does not allow us to predict risk for psychopathology in a specific manner, perhaps P3AR combined with a behavioral indicator of impaired inhibitory control could provide stronger evidence of an individual’s “true” level of externalizing. Furthermore, in the present study the combination of P3AR and inhibitory control did not predict internalizing symptoms, supporting the idea that these measures demonstrate specificity to externalizing.

Researchers have explored the validity of a multivariate psychophysiological composite as an index of proneness to externalizing psychopathology [97] and have argued for the value of using psychophysiology to tap externalizing-relevant neurobehavioral traits, namely inhibitory control [98]. However, it is unclear why a more direct behavioral measure of inhibitory control would not be equally, if not more, useful for quantifying inhibitory control deficits, particularly if multiple behavioral indices are combined (see [99]). Based on the present results, it is reasonable to postulate that a composite including behavioral measures in addition to biological (and self-report) measures could incrementally improve the validity, specificity, and reliability of an index of risk for externalizing psychopathology. A range of measurement domains (neural, behavioral, self-report) was included in a “psychoneurometric” approach to assessing internalizing psychopathology detailed by Moser et al. [100], and this same multi-domain approach could strengthen the assessment of externalizing as well. A multivariate composite consisting of measures from multiple domains would also alleviate concerns about the issue of method variance in the quantification of risk [98], as it would help ensure the aggregation of variance shared due to underlying processes rather than simply the same type of measurement tool. Overall, the quantification of externalizing risk stands to benefit from integrating mechanistically informative behavioral measures with biological measures.

An additional way in which P3AR has fallen short of its initial promise is its failure to deliver the “main advantage” of an endophenotype: facilitating the discovery of susceptibility genes (1). In the first published genome-wide association study of P3, Malone et al. [101] measured P3 during a standard oddball paradigm and found that P3 amplitude was significantly associated with only one out of hundreds of thousands of genetic variants. Reflecting on this disappointing outcome, Iacono et al., 2014 [102] were led to question the utility of endophenotypes and conclude that there is “little reason to expect further refinement of the endophenotype to lead to valid genetic associations” (p. 1344). Yet, there may be reason to believe that this conclusion is premature. Given that there has long been a disconnect between conceptualizations of cognitive deficits in externalizing and the psychobiological measures (i.e., P3AR during oddball) used to quantify these deficits, an enriched conceptualization of externalizing risk incorporating brain-behavior relationships could prove advantageous. Considering P3AR alone likely introduces considerable “noise” in genetic association studies, drowning out the true associations between P3AR and externalizing-related genes. In contrast, considering P3AR in concert with behavioral measures of inhibitory control might increase specificity so that genes can be linked to cases at risk for externalizing outcomes, and not “false positives” (e.g., someone with P3AR who is actually at risk for psychosis rather than an externalizing disorder). Taken together, perhaps the key challenge in harnessing the utility of endophenotypes (and biological measures in general) is to bridge levels of analysis, linking brain and behavior in a mutually explanatory manner to arrive at multivariate endophenotypes.

4.3. Conclusions

Overall, the present results provide direct support for the contention that P3AR reflects executive dysfunction in externalizing. Given that externalizing psychopathology creates an enormous burden on society—for victims of crime, family members of individuals with addiction, and individuals who struggle to regulate their behavior—understanding the processes that contribute to risk for externalizing disorders is critical. The application of biological measures represents a valuable frontier in terms of providing objective reflections of externalizing-relevant constructs. At the same time, behavior measured under precisely specified and controlled conditions often can yield insight into key processes underlying externalizing. Considering psychobiological measures alongside behavioral measures, in certain contexts, may represent meaningful progress toward better conceptualizing underlying processes and ultimately can strengthen our efforts to quantify risk and understand the complex, multi-faceted etiology of externalizing psychopathology.

Acknowledgments

We sincerely thank Linda Mayes, Michael Crowley, and Jia Wu for their assistance with this research.

References


