Archival Report

Biological Psychiatry: CNNI

Internal Consistency of Functional Magnetic Resonance Imaging and Electroencephalography Measures of Reward in Late Childhood and Early Adolescence

Katherine R. Luking, Brady D. Nelson, Zachary P. Infantolino, Colin L. Sauder, and Greg Hajcak

ABSTRACT

BACKGROUND: Abnormal neural response to reward is increasingly thought to function as a biological correlate of emerging psychopathology during adolescence. However, this view assumes that such responses have good psychometric properties, especially internal consistency—an assumption that is rarely tested.

METHODS: Internal consistency (i.e., split-half reliability) was calculated for event-related potential (ERP) and blood oxygen level–dependent (BOLD) responses to monetary gain and loss feedback from the same sample of 8- to 14-year-old girls ($N = 177$). Internal consistency for ERP (i.e., feedback negativity) and BOLD responses within the ventral striatum and medial and/or lateral prefrontal cortex to gain, loss, difference scores (gain − loss), and residual scores (gain controlling for loss) was compared. Moderation analyses were conducted to investigate whether internal consistency differed by age.

RESULTS: ERP and BOLD responses to gain and loss feedback showed high internal consistency in all regions (Spearman-Brown coefficients $\geq .70$). When considering difference and residual scores, however, responses showed lower internal consistency (Spearman-Brown coefficients $\leq .50$), with particularly low internal consistency for subtraction-based scores (Spearman-Brown coefficients $\leq .36$). Age was not a significant moderator of split-half relationships, indicating similar internal consistency across late childhood to early adolescence.

CONCLUSIONS: Within the same subjects, high internal consistency was observed for both ERP and BOLD responses to gains and losses, which did not vary as a function of age. Moreover, excellent psychometric properties were evident even within the first half of the experiment. Difference scores were characterized by lower internal consistency, although regression-based approaches outperformed subtraction-based difference scores.

Keywords: Adolescence, Childhood, ERP, fMRI, Internal consistency, Reward

http://dx.doi.org/10.1016/j.bpsc.2016.12.004

In the past decade, efforts to leverage neuroscience-based measures to shed light on psychopathology have increased. For example, the National Institute of Mental Health Research Domain Criteria initiative aims to move psychiatric research toward investigating neurobiological mechanisms of specific domains of function (1). Such biologically oriented approaches to understanding psychopathology and individual differences require neural measures with good psychometric properties (2). No measure can be valid if it is not reliable (3,4). Yet few studies have considered measurement properties of neural metrics. In this article, we focus on internal consistency reliability (referred to as internal consistency throughout) of the neural response to reward feedback (both reward gains and losses) given its central role in the "positive valence systems" of the Research Domain Criteria matrix and considering that abnormal response to reward is a promising biomarker of risk for multiple types of psychopathology (5,6).

Establishing whether neural response to reward feedback shows adequate internal consistency is a critical first step in determining the utility of such individual difference measures. Neural responses to gains and losses show dramatic changes over the course of adolescence (7–9) with abnormal responses to reward prospectively predicting increases in depressive symptoms and substance use during adolescence (5,6,10,11). Such findings have led to the idea that abnormal neural response to reward may confer vulnerability to depression and other disorders during development, when patterns of normative reward response are in flux (7). As such, investigating internal consistency of neural reward response during emerging adolescence, a time of increasing onset of reward-related

ISSN: 2451-9022

psychopathology particularly for female subjects (12), is the aim of the current study.

Several studies have examined test-retest reliability of neural measures of reward response [e.g., (13–18)]. Test-retest reliability refers to the consistency of individuals' scores assessed at different points in time. To assess test-retest reliability of blood oxygen level–dependent (BOLD) responses and event-related potentials (ERPs), participants complete an identical task in two sessions separated by weeks or months. Neural responses are then correlated across sessions. High test-retest reliability indicates relative stability of scores over time and suggests that a measure has trait-like properties. In the domain of reward, ventral striatal (VS) response to reward feedback has been found to have relatively low test-retest reliability in some studies (13), whereas other studies report moderate test-retest reliability (14,15). ERPs indexing reward response, including feedback negativity and reward positivity (19,20), tend to show higher test-retest reliability (16–18)—although these measures have never been directly compared in the same subjects. It is important to note that test-retest reliability is sensitive to between-session variations in methodological noise (e.g., thermal and system noise and motion artifacts in functional magnetic resonance imaging [fMRI]) and participant state (e.g., time of day, fatigue, or stress level) (21). Test-retest reliability does not address whether a measure is internally consistent. It is possible for a measure to be internally consistent despite low test-retest reliability (i.e., if a measure was highly sensitive to state-related variables); moreover, it is possible for a measure to be reliable over multiple administrations, despite having poor internal consistency (i.e., if a measure comprised multiple trait-like items that were themselves unrelated to one another).

Internal consistency refers to the similarity of items on a measure. A relatively simple method of calculating internal consistency is the split-half method, which involves correlating two separate scores for each subject that are derived from the same measure at the same testing session. Most frequently, scores on odd and even trials are used. Because splitting the data artificially reduces the number of trials by half, Pearson coefficients (r) are corrected using the Spearman-Brown coefficient (SB) prediction formula: $SB = 2r/(1 + r)$.

Internal consistency is a fundamental measurement property that places an upper limit on how well a measure can index individual differences. A measure cannot correlate better with another measure than it correlates with itself; thus, poor internal consistency will limit the ability of a measure to relate to other individual difference variables. This point has been central to several debates regarding the reproducibility and implications of studies linking BOLD response to individual differences (21–25). However, the internal consistency of the BOLD response to reward has not been examined to date; in fact, internal consistency of fMRI measures is hardly ever examined and reported in fMRI studies. This is surprising given the increasing focus on how VS reward response relates to individual differences in symptom severity and risk factors for psychopathology in both adults [e.g., (26,27)] and adolescents [e.g., (28–32)]. However, reward-related ERPs are characterized by high split-half internal consistency even during late childhood and early adolescence (17,18,33), although it is unclear whether internal consistency changes as a function of age during adolescence. Moreover, no studies have directly compared psychometric properties of ERP and fMRI measures of reward in the same subjects.

We assessed internal consistency of both ERP and BOLD responses to reward gain and loss feedback in a large unselected sample of girls during emerging adolescence. Furthermore, we tested the hypothesis that internal consistency will be comparable across age during emerging adolescence despite the normative increases in response to reward feedback observed during these ages [e.g., (8)]. This knowledge is critical not only for determining the utility of VS and/or ERP response to reward as an early biomarker of risk but also for interpreting developmental changes in reward-related responses. That is, presumed "developmental" differences in reward-related neural responses could be attributed to variation in internal consistency as a function of age. We also compare the internal consistency of BOLD and ERP responses to reward within the same subjects to examine potential methodological differences in assessing reward-related neural activity. Finally, as internal consistency may differ as a function of task length, particularly when a limited number of stimuli are repeated many times over a longer task, we report internal consistency of early and late components of the task. Investigating these properties of internal consistency during emerging adolescence will provide a critical baseline for future inquiries regarding relationships between neural reward response and psychopathology risk during this time of increasing vulnerability to psychopathology.

METHODS AND MATERIALS

Participants

fMRI scanning was performed in 198 girls and their parents, from a parent study [described in (34)]. This study focused on neural response to reward in girls during emerging adolescence, as studies in the literature on depression risk are increasingly focusing on female subjects, given the increased incidence of depression in women, a sex difference that emerges over adolescence (35). Of 198 girls, 177 provided sufficient quality data from both fMRI and electroencephalography (EEG) versions of the doors guessing task and were included in the current study (5 were excluded for excessive motion, and 16 were excluded for scanner sequence or other mechanical error). Girls were 8 to 14 years old (mean, 12.56 years; SD 1.80); 77.4% were Caucasian, 16.9% were African American, 2.3% were Hispanic, and 3.4% were identified as "other." Informed assent and consent were obtained from the participants and their parents, respectively, before participation. The Stony Brook University Institutional Review Board approved the research protocol.

Measures

The doors task was similar to versions used in previous studies (36). Participants were presented with two identical doors and instructed to select the left or right door via mouse click. They were told that they would gain $0.50 or lose $0.25 on each trial depending on whether the "correct" door was selected. The order of gain and loss feedback events was predetermined such that all participants experienced 30 gain feedback events and 30 loss feedback events presented in a pseudorandom order.