Temperament and Psychopathology: The "Community" To Which You Belong Matters

Abstract

We utilized a community detection approach to longitudinally (a) identify distinct groups of children with common temperament profiles in infancy and at 2 and 3 years of age and (b) determine whether co-occurrence of certain temperament traits may be early predictors of internalizing problems at 5 years of age. 774 infants (360 girls; 88.6% White, 9.8% Hispanic, and 1.6% other races) were recruited from the Boston area. Data collection spanned from 2012 to 2021. The analysis yielded three distinct groups of children with different temperament traits and were associated with significant variation in levels of internalizing symptoms and anxiety diagnosis rate. Our findings suggest that stable temperament "communities" can be detected in early childhood and may predict risk for psychopathology later in life.

Keywords: *child temperament, community detection, psychopathology, anxiety*

Introduction

Temperament profile refers to a relatively stable disposition (e.g., rank-order stability) accompanied by mean-level changes over time – it emerges early in development and underlies and modulates the expression of activity, reactivity, emotionality, and sociability (Stifter & Dollar, 2016). Theoretical models of temperament converge on three main components: negativity, surgency, and effortful control (Rothbart et al., 2000). These components appear stable in childhood and have been replicated in different cultural contexts (Ahadi et al., 1993). Taking a dimensional approach, each of these high-order components comprises different aspects of behavior or temperament traits, with negativity including the expression of negative emotions such as anger, fear, and sadness; surgency including approach behaviors, activity level, and impulsivity; and effortful control encapsulating the ability to control attention and behavior (Rothbart & Bates, 2007). While the constituent dimensions of temperament are largely agreed upon (Shiner et al., 2012), there is not yet a consensus on how individual differences in the constellation of temperament traits ought to be characterized (Fu & Pérez-Edgar, 2015; Ostlund et al., 2021a). Distinct from theories holding a dimensional view of temperament (e.g., Goldsmith, 1996; Rothbart et al., 2000), Kagan and colleagues defined temperament categories and emphasized the use of behavioral observations rather than parent-report questionnaires to study temperament (Kagan, 2003). For instance, behavioral inhibition (BI) is a temperament category characterized by discrete biological dispositions marked by elevated psychophysiological activity, hyper-vigilance, and behavioral withdrawal upon encountering novel people and objects (Kagan et al., 1984).

Person-centered data-driven approaches, such as clustering analysis and latent profile analysis (LPA), have the unique advantage of leveraging the strengths of both dimensional and

categorical perspectives on child temperament. These approaches have been used to identify categories of temperament by grouping together individuals with similar configurations of temperament traits that are individually dimensional (Bergman & Magnusson, 1997). Underlying these approaches is the assumption that the complex interplay among multiple dimensionally based temperament-relevant behaviors gives rise to certain patterns of trait expression that are shared among the same categories of children (Ostlund et al., 2021b). Person-centered approaches also provide increased statistical power, relative to individual correlations, and open the possibility of individual prediction based upon a broad temperament profile rather than a single dimension (Stifter & Dollar, 2016).

Three to four temperament profiles have frequently been identified in prior studies using person-centered data-driven approaches. One profile is characterized by high negative affect and below average regulatory abilities, which has been named with terms such as “negative reactive,” “dysregulated,” “unregulated,” and “undercontrolled” by different authors (Beekman et al., 2015; Komsi et al., 2006; Lin et al., 2018; Prokasky et al., 2017; Scott et al., 2016). A second often-identified profile in these studies shows opposite temperament traits compared to the first profile, characterized by high levels of positive affect and regulation and often named with terms such as “positive affect/reactive,” “well regulated,” and “resilient.” A third profile that has been identified is characterized by extremely low level of activity and moderate to high level of negative affect (e.g., fear), which has been labeled “overcontrolled” (Komsi et al., 2006) and “fealful” (Beekman et al., 2015) in different studies. Yet another profile is characterized by high activity level and above-average negative affect and has been labeled “high reactive” (Prokasky et al., 2017) and “active reactive” (Beekman et al., 2015). Similar profiles have been found with different methods (e.g., k-means clustering vs. LPA; Gartstein et al., 2017) and in

children with different racial background and risk status (e.g., attention deficits, prenatal substance exposure) (Karalunas et al., 2014; Lin et al., 2018, 2021). For example, a recent study found three profiles (i.e., high positive affect and well-regulated, negative reactive and low-regulated, and low positive affect and low-regulated) in Mexican-American children (Lin et al., 2021).

There is mounting evidence to support an association between certain temperamental profiles that emerge in early childhood and internalizing disorders (e.g., anxiety diagnoses) that manifest in preschool age through adolescence (Pérez-Edgar & Fox, 2005). Internalizing disorders are marked by measurable symptoms of anxiety, fearfulness, social withdrawal, and/or depressed mood (Kamphaus & Mays, 2011). Early childhood is a critical period for the development of internalizing disorders that have negative effects on academic, social, and adaptive functioning throughout life (Beesdo et al., 2009). Behaviorally inhibited children are often reticent in unfamiliar social situations and have been shown to be at elevated risk for developing anxiety and internalizing problems (Kagan & Snidman, 1999; LoBue & Pérez-Edgar, 2014). Person-centered approaches have also been used to study the link between temperament profiles and childhood psychopathology. For example, stable high versus stable low profiles have been found to emerge across infancy and early childhood for BI and exuberance (Putman & Stifter, 2005), and these longitudinal profiles have been associated with later social anxiety disorders (Chronis-Tusano et al., 2009; Degnan et al., 2010).

Distinct from the research on BI and social anxiety, the majority of studies on the relation between temperament and psychopathology have regarded temperament factors as separate dimensions. For example, high negative emotionality has been linked with increased risk of internalizing and externalizing problems across childhood (Gartstein et al., 2012). In contrast,

high effortful control has been found to be a protective factor associated with lower levels of internalizing and externalizing problems (Eisenberg et al., 2010). However, as higher-level factor structure analyses show, these temperament dimensions do not exist in isolation but often co-occur in consistent configurations within individuals (Gartstein et al., 2017). Thus, a certain pattern of co-occurrence of facets of temperament may better capture the phenotypic heterogeneity inherent in aberrant behaviors and predict risk for internalizing disorders than a particular temperament trait/dimension (Ostlund et al., 2021a). To this end, using advanced data-driven approaches to extract homogenous phenotypes and profiles of traits related to psychopathology is promoted by the Research Domain Criteria (RDoC) framework (Cuthbert, 2014).

Current study

The developmental trajectory of temperament profiles in early childhood is not yet fully understood, and most research in this area has examined the association of temperament profiles with laboratory performance, cognitive functions, and psychopathology separately. There is a need for a systematic longitudinal examination of the developmental trajectory of temperament profiles over early childhood, how such profiles are associated with internal cognitive functions measured with laboratory observations of behavior, and how such configurations predict later internalizing problems. The present study addressed this gap by leveraging a large longitudinal sample to apply person-centered computational approaches to detect temperament communities (i.e., groups) over the first 3 years of life. We employed a data-driven clustering algorithm that makes no assumptions about the size and shape of potential clusters in comparison to commonly used clustering methods, like k-means clustering or LPA. The only assumption is that individuals within a cluster are more similar to each other than they are to individuals in other clusters. We

hypothesized that the current analysis would generate 3 to 4 temperament groups that share features with profiles detected in previous studies. We further hypothesized that the same groups will be detected across ages at the group level, but that individual group membership will show some variability across time. The current analysis is relatively exploratory, as no prior study has used the current clustering algorithm to detect temperament group in early childhood.

To externally validate the derived temperament groups, we examined children's behavioral performance in an emotion disengagement task, a prosocial task, and a BI task, as studies have shown associations between children's temperament traits and their behaviors in these experimental tasks. For example, greater difficulty in disengaging from fearful faces has been observed in 12-month-old infants with higher negative affectivity (Nakagawa & Sukigara, 2012). Prosocial behavior has been found to relate positively to sociability and activity and negatively to shyness and negative emotionality in children and adolescents (e.g., Carlo et al., 2012; Gross et al., 2015). In addition, children with a combination of low levels of self-regulation and high levels of negative emotionality at 54 months have been shown to be less prosocial in later childhood than children with other temperament profiles (Laible et al., 2014). Moreover, children rated as high on BI on parent-report questionnaires are more likely to show the typical BI behaviors in laboratory settings, e.g., longer latency to initiate contact with a stranger (Bishop, Spence, & McDonald, 2003). Of specific interest is to examine how these behaviors might differ among temperament groups detected by our data-driven approach.

Further, we examined how these temperament groups were associated with internalizing symptoms and anxiety symptoms and diagnoses at age 5 years. Specifically, we assessed whether temperament groups detected by clustering analysis across infancy to 3 years were associated with differences in internalizing symptoms generally and anxiety symptoms

specifically and with differential risk for an anxiety diagnosis at age 5 years. We also investigated whether the trajectory of temperament profile, e.g., children stayed in a certain group vs. children who transitioned to different groups, is associated with the clinical outcomes. Finally, we tested whether any findings were specific to internalizing problems by examining relations between temperament groups and externalizing problems.

Method

Participants

Participants were recruited from a registry of local births comprising families who had indicated willingness to participate in developmental research from the Boston area. Families for the current analyses participated in a prospective study (N=774) to examine the early development of emotion processing. Children were recruited at ages 5 months, 7 months, or 12 months (T1: Time 1; questionnaires, laboratory battery) and followed when the child was ages 2 years (T2: Time 2; questionnaires), 3 years (T3: Time 3; questionnaires, laboratory battery), and 5 years (T4: Time 4; questionnaires, clinical interview, laboratory battery). Exclusion criteria are described in the Supplementary Information (SI) document. The vast majority of the participants were White (88.6%); 9.8% identified as Hispanic, including 2.5% Mexican American, 1.3% Puerto Rican, 0.4% Cuban, 5.0% other Hispanic and Latino/a, 0.6% mixed Hispanic and Latino/a, and 1.6% not reported. At the time of T1 visit, which occurred between 2012 and 2017, more than half of the families (59.4%) reported an annual household income greater than $50,000, and 62.7% of mothers and 56.7% of fathers had completed a master's degree or higher. Details on the sample sociodemographic characteristics are provide in SI Table 1. The current analytic sample included 700 children with valid temperament data at the infancy (T1) timepoint (5 months: N = 190, 7 months: N = 213, or 12 months: N = 297). No differences in

sociodemographic data were found between participants with and without T2 and T3 temperament data (Behrendt et al., 2020). The number of children with temperament data by age and sex can be found in Table 1.

[Table 1 goes here]

Ethics statement

Parents of the participants provided written informed consent before each of the child's study visits, and ethical permission for the study was obtained from the Institutional Review Board at Boston Children's Hospital.

Materials and Methods

Child temperament

Child temperament was assessed at T1, T2, and T3. At T1, mothers completed the Infant Behavior Questionnaire-Revised (IBQ-R; (Putnam et al., 2014)) prior to the T1 laboratory visit. The IBQ-R comprises 14 subscales, which factor analyses show contribute to composite measures for three dimensions of child temperament: surgency/extraversion (Cronbach's α = .814), negative affectivity (α = .798), and orienting/regulation (Cronbach's α = .695). At T2 and T3, mothers completed the Early Childhood Behavior Questionnaire (ECBQ; Putnam et al., 2006). The ECBQ also provides three composite measures of surgency/extraversion (T2 α = .698, T3 α = .708), negative affectivity (T2 α = .797, T3 α = .766), and effortful control (T2 α = .767, T3 α = .798), an age-upward extension of the IBQ-R orienting/regulation factor. The subscales for IBQ-R and ECBQ have been listed in Fig 1,2 and described in the SI document.

For both the IBQ-R and ECBQ, mothers rated the frequency that their child engaged in specific day-to-day behaviors in the prior one to two weeks using a 7-point scale, with responses ranging from 1 (never) to 7 (always). Item scores were summed and averaged according to

measure scoring rules to create subscale scores, with higher scores indicating greater levels of that temperament dimension. Most of the participants with temperament data had no missing subscales (i.e., had complete data): 677 / 700 in infancy, 509 / 516 at T2, and 463 / 467 at T3. The participants with missing data only missed one or two subscales, likely due to a mistake when the mother completed the questionnaire, or an error occurred during data input. Missing data were imputed with the Multiple Imputation by Chained Equations (MICE) in Python.

Internalizing symptoms and anxiety symptoms and diagnoses

At T4, mothers completed the Child Behavior Checklist 1½ -5 (CBCL/1½ -5 (Achenbach & Rescorla, 2000)). The CBCL is one of the most well-established, empirically supported questionnaires to assess child psychopathology symptoms (Achenbach et al., 2008). It produces scores on multiple syndrome and DSM-oriented scales as well as higher-order symptom scores. For the current analyses, the composite Internalizing Problems scale and the specific DSM-oriented Anxiety Problems scale were used. The Internalizing Problems score is composed of the following syndrome scales (α = .776): Anxious/Depressed, Emotionally Reactive, Withdrawn, and Somatic Complaints. The Anxiety Problems scale is specifically aligned with Diagnostic and Statistical Manual (DSM-5) anxiety disorders. Raw scores for each scale score were transformed to T scores.

At T4, mothers were also invited to complete the Diagnostic Infant and Preschool Assessment (DIPA), a semi-structured clinical interview for caregivers of young children that was administered to obtain ratings of lifetime and current psychiatric disorders (Scheeringa & Haslett, 2010). The DIPA has demonstrated reliability and validity in assessing clinical symptoms in research with very young children and was developed such that it can be administered by trained research staff without specific clinical mental health experience/training.

The DIPA was administered as an interview by trained research staff, supervised by a licensed clinical psychologist (author MBE). The version of the DIPA administered in the current study (version 7/12/14) assesses a variety of DSM-5 disorders and other problems, including anxiety disorders, mood disorders, and externalizing disorders (Attention Deficit/Hyperactivity Disorder [ADHD], Oppositional Defiant Disorder [ODD], Conduct Disorder [CD]), each in a self-contained module. For the current analyses, children were coded as to whether they currently (i.e., at age 5 years) met criteria (0: no diagnosis; 1: one or more diagnoses) for at least one of the following anxiety disorders: Generalized Anxiety Disorder, Social Anxiety Disorder (Social Phobia), Separation Anxiety Disorder.

Experimental tasks – Prosocial, emotion disengagement, and BI tasks

Children completed a prosocial task during their T3 lab visit. This task included the "Bin" and "Clothespin" episodes adapted from (Warneken & Tomasello, 2006). In these tasks were observed for whether they would help others to achieve their goals. Children who helped before being explicitly asked were categorized as "spontaneous helpers"; those who helped after being asked were categorized as "prompted helpers"; those who did not help were categorized as "non-helpers."

An eye-tracking task designed to assess attention disengagement was administered at visit T4 (Xie et al., 2021). A stimulus was first presented on the center of the screen for 4000 ms. This center stimulus was randomly chosen from four types of images: a non-face pattern, and angry, fearful, or happy faces. A second (target) stimulus was presented with a 200 ms onset asynchrony laterally on the left or right side of the screen with 13.6° eccentricity and remained on the screen for 300 ms. The dwell time variable was extracted from the eye-tracking data,

calculated as the normalized duration of fixation on the central face before the saccade to the lateral target between 150 ms and 1000 ms following the target onset.

In order to measure BI in the laboratory setting, children at T3 were asked to complete a series of tasks adopted from Fox et al. (2001). Children's behavior in response to three unfamiliar stimuli was coded. In the first task, a "stranger" sat in silence with her head down for one minute before taking out a toy truck and playing with it for another minute in silence. The stranger continued to play with the truck for one minute and then invited the child to play with them twice during that time if the child did not voluntarily approach the stranger to play. During this time, the mother was seated in a corner of the room, wearing a pair of headphones while filling out a set of questionnaires. Children's 1) latency to touch the toy truck and 2) time spent in proximity (i.e., a 1-ft. radius) to the mother while the stranger was present were coded and transformed to standardized z-scores. In the other tasks, a toy robot and a tunnel were used (i.e., child invited to play with robot and climb through tunnel), each with their two measures (latency and time z-scores) calculated. A BI composite score was created by summing the z-scores. Please see the SI document for greater detail for these behavioral tasks.

Clustering analysis of temperament groups

We employed an optimization clustering method, i.e., community detection, to develop temperament groups based on the temperament data collected at T1, T2, and T3. Unlike other methods commonly used in the literature, like k-means clustering or latent profile analysis, community detection does not require the researcher to make assumptions about the shape and size of the clusters that are difficult to justify, e.g., assuming similar distributions of the clusters in multidimensional space. The similarity of participants was expressed as the Pearson correlation between their IBQ-R or ECBQ scores. The objective of the community detection was

to find a subdivision of the participant-by-participant correlation matrix with the maximum correlation within each subgroup and minimum correlation between the subgroups. To this end, the Louvain algorithm was used for community detection, followed by a fine-tuning step using the Keringhan-Lin algorithm. We employed an adapted implementation of the Louvain algorithm that can incorporate signed distances (Rubinov & Sporns, 2011), i.e., negative correlation values. The algorithm was run 100 times to construct an agreement matrix, which was then used to obtain a consensus community partition. We repeated this procedure for multiple resolutions (varying γ between .1 and 5.0). More details about the method and the codes can be found in the SI document.

Results

Section 1. Temperament groups

The descriptive statistics for temperamental dimensions and for their subscales, by age and by sex, can be found in Table 1 and SI Table 2, respectively. The p-values for post-hoc comparisons reported in the following sections have been adjusted with a false discovery rate (FDR) of .05 or Bonferroni correction.

Infant (IBQ-R) Results (T1)

There were significant differences in temperament scores between the age groups in the first year of life (5, 7, 12 months, see SI for a detailed analysis). Because the clustering analysis is more powerful with larger sample sizes, we elected to regress out the age effect rather than conduct the analysis separately by infant age. Male and female infants were pooled together in these analyses because no sex effects were found (SI Table 2).

The clustering analysis indicated the presence of three temperament groups in infancy (see Fig 1 for characteristics by age and by sex and Fig 2 for the Silhouette plot). The groups

displayed contrasting temperament profiles (ANOVA: scale x group interaction: $F(26, 8684) = 77.07$, $p < .001$, $\eta^2 = .084$). One group (C1: N = 293) was characterized by high ratings of surgency/extraversion, low ratings of negative affectivity (note: falling reactivity has negative loadings on negative affectivity), and high ratings of orienting/regulation. A second group (C2: N = 304) showed an opposite pattern, i.e., moderate levels of surgency/extraversion, high ratings of negative affectivity, and low ratings of orienting/regulation. A third group (C3: N = 103) displayed low levels of surgency/extraversion, moderate levels of negative affectivity, and mixed ratings on the subscales of orienting/regulation, specifically low ratings on low pleasure and duration of orienting but high ratings on cuddliness and soothability.

Based on the features of each group, we labeled them as follows: C1 – "Emotionally and Behaviorally Regulated (EBR)", C2 – "Emotionally and Behaviorally Dysregulated (EBD)" and C3 – "Introverted and Over-controlled (IOC)." The clustering analysis was additionally run using the complete data (N = 677), and the same temperament groups were detected (SI Fig 1A).

[Figure 1 goes here]

[Figure 2 goes here]

Two- and Three-Year (ECBQ) Results (T2, T3)

Clustering analysis was conducted separately for boys and girls because of significant differences in temperament scores (see SI). The clustering analysis of the T2 ECBQ data showed similar results for boys and girls, and similar results were obtained when the data were pooled together (SI Fig 1B). The analysis indicated the presence of two larger groups and one smaller group (Fig 1) that differed in temperament traits, boys: scale x group: $F(34, 4753)=36.79$, $p < .001$, $\eta^2 = .081$; girls: scale x group: $F(34, 3978)=24.50$, $p < .001$, $\eta^2 = .073$. These groups showed characteristics that were comparable to those detected at T1. Specifically, one group (N

= 195) showed high ratings on surgency/extraversion and effortful control and low ratings on negative affectivity (note: Soothability has a negative loading on negative affectivity), and thus was referred to as EBR. A second group (N = 187), referred to as EBD, was characterized by the opposite pattern. A third group (N = 134), referred to as IOC, featured low surgency/extraversion and relatively high loadings on a few of the subscales of negative affectivity and all subscales of effortful control. The children in this IOC group also had low ratings on motor activation and frustration and high ratings on soothability.

These groupings were stable across the T2 and T3 assessments, i.e. the characterization of the temperament groups detected at T3 was remarkably consistent with that detected at T2, with minor differences between two time points (Fig 1). Thus, the same labels were used to characterize the groups at T3, i.e., the EBR (N = 177), EBD (N = 178), and IOC (N = 112) groups. Please see the SI for details.

Stability of individual temperament group membership across age

Transitions between temperament groups were assessed separately from T1 to T2 and from T2 and T3. Of the 470 children with valid temperament data at both T1 and T2, approximately half (47.7%) remained in the same group across assessments. The stability of group membership increased with age, as 62.4% of the 367 children with T2 and T3 data remained in the same group; T2 to T3 (47.7%) vs. T1 to T2 (62.4%), $\chi^2(1, N = 837) = 18.03$, $p < .001$, $\Phi = .623$ (Fig 2). Moreover, children in the EBR and EBD groups were less likely to transition to other groups compared to those in the IOC group (Fig 2). Please see SI for details.

External validation: Temperament groups show differences in eye-tracking, prosocial and BI behaviors

A mixed ANOVA was run to determine whether children's looking behavior in the eye-tracking task was different between temperament groups derived at T3. In this model, temperament group was included as a between-subject factor (3 levels) and stimulus type (4 levels) was included as a within-subject factor (4 levels: angry, fearful, happy faces and the non-face geometric pattern). The analysis revealed significant main effects of temperament group ($F(2,185) = 3.91$, $p = .022$, $\eta^2 = .042$) and stimulus type ($F(3,555) = 10.24$, $p < .001$, $\eta^2 = .055$), but no interaction was found between the two factors ($F(3,555) = 2.00$, $p = .113$, $\eta^2 = .022$). Post-hoc comparisons with a Bonferroni adjustment revealed that the DT on the stimuli was longer for children in the EBR group compared to children in the IOC group, $M_{diff} = .074$, 95%CI = [.0030 .15] (Fig 3A). Again, the post-hoc comparisons reported in this section had been adjusted with a false discovery rate (FDR) of .05.

Children's prosocial behaviors were compared between different temperament groups to further validate the clustering analysis results. The percentage of "spontaneous helper," "prompted helper," and "non-helper" in each temperament group was compared via chi-square tests. These analyses revealed that children in the IOC group were less likely to be a "spontaneous helper" and more likely to be a "non-helper" in the Bin episode compared to children in the other two groups: for "spontaneous helper," IOC vs. EBR, $\chi^2(1, N = 216) = 5.86$, $p = .016$, $\Phi = .399$; IOC vs. EBD, $\chi^2(1, N = 210) = 5.86$, $p = .039$, $\Phi = .404$; for "non-helper," IOC vs. EBR, $\chi^2(1, N = 216) = 7.07$, $p = .008$, $\Phi = .481$; IOC vs. EBD, $\chi^2(1, N = 210) = 4.57$, $p = .032$, $\Phi = .315$ (Fig 3B). No difference was found among the temperament groups for the clothespin episode (SI Fig 3).

Children's BI composite scores were compared among the temperament groups via Mann-Whitney U tests. This analysis revealed that the BI composite scores of the IOC children

(*Mdn* = 1.518, N = 58) were higher than those of the EBR (*Mdn* = -.881, N = 97) and EBD (*Mdn* = -.542, N = 90) children, IOC vs. EBR: $U = 2252.0$, z = 2.081, $p = .037$, $\eta^2 = .0281$; IOC vs. EBD: $U = 2037.0$, z = 2.253, $p = .024$, $\eta^2 = .035$ (Fig 3C). No difference was found between the EBR and EBD groups. Non-parametric tests were conducted because the BI composite score was not normally distributed as shown by the Kolmogorov-Smirnov (KS) tests (EBR, KS distance = 0.111, $p = .005$; EBD, KS distance = .097, $p = .038$, IOC, KS distance = 0.138, $p = .008$).

[Figure 3 goes here]

Section 2. Relation between temperament profiles and psychopathology

Temperament groups and internalizing symptoms

One-way ANOVAs were conducted to determine whether internalizing symptoms at age 5 years differed by temperament group at individual ages (T1, T2, T3). The analyses revealed main effects of temperament group on 5-year CBCL Internalizing Problems scores at each age temperament was assessed: T1 $F(2, 338) = 5.23$, $p = .0058$, $\eta^2 = .030$; T2 $F(2, 279) = 15.80$, $p < .001$, $\eta^2 = .102$; T3 $F(2, 271) = 20.02$, $p < .001$, $\eta^2 = .040$ (Fig 4A). Post-hoc multiple comparisons showed that, compared to children belonging to the EBR group at T1 (i.e., during infancy), children belonging to the EBD group at T1 showed greater internalizing symptoms at 5 years of age ($M_{diff} = 3.51$, 95%CI = [0.66 6.35]). Children in the EBD group at T2 showed greater internalizing symptoms than children in the other two groups at T2 (EBD vs. EBR: $M_{diff} = 6.88$, 95%CI = [3.83 9.94]; EBD vs. IOC: $M_{diff} = 5.68$, 95%CI = [2.00 9.37]). Children in the EBD and IOC groups at T3 showed greater internalizing symptoms than children in the EBR group at T3 (EBD vs. EBR: $M_{diff} = 7.99$, 95%CI = [4.90 11.08]; IOC vs. EBR: $M_{diff} = 5.56$, 95%CI = [2.09 9.03]); no difference was found between children in the EBD and IOC groups at T3 (Fig 4A).

The same analyses were applied to test the effects of temperament group on CBCL Anxiety Problems scores. There was a main effect of temperament group at all three ages on anxiety symptoms at age 5 years: T1, $F(2, 338) = 4.19$, $p = .016$, $\eta^2 = .0242$; T2, $F(2, 279) = 5.33$, $p = .005$, $\eta^2 = .037$; T3, $F(2, 271) = 4.99$, $p = .007$, $\eta^2 = .024$ (Fig 4B). Post-hoc comparisons showed higher 5-year anxiety scores among children in the EBD compared to children in the EBR groups for all three ages, T1: $M_{diff} = 1.96$, 95%CI = [.22 3.69]; T2: $M_{diff} = 2.35$, 95%CI = [.53 4.17]); and T3: $M_{diff} = 2.30$, 95%CI = [.19 4.40]). Children in the IOC group at T2 and those in the IOC group at T3 also showed higher anxiety scores compared to children in the EBR group at the same age, T2: $M_{diff} = 2.13$, 95%CI = [.017 4.27]); T3: $M_{diff} = 2.72$, 95%CI = [.35 5.09]) (Fig 4B).

[Figure 4 goes here]

Temperament groups and anxiety diagnoses

The probability of meeting criteria for one or more anxiety diagnoses at age 5 years was compared among the temperament groups at different ages using chi-square tests. These analyses showed that the association between temperament profile and anxiety diagnosis strengthened with age. Specifically, the proportion of children having an anxiety diagnosis at age 5 years did not differ by temperament group at T1: EBR, 8.7%; EBD, 14.2%; IOC, 10.5% (Fig 5). The proportion of children having an anxiety diagnosis at age 5 years differed by temperament group at T2: More children in the EBD group (17.6%) had an anxiety diagnosis than those in the EBR group (5%), $\chi^2(1, N = 186) = 7.75$, $p = .005$, $\Phi = .568$; neither group differed from the IOC group (10.2%) (Fig 5). Children who were in either the EBD (16.5%) or the IOC (11.5%) group at T3 were more likely to have an anxiety disorder at age 5 years compared to children in the

EBR group (2.2%): EBD vs. EBR, $\chi^2(1, N = 181) = 10.81$, $p = .001$, $\Phi = .793$; IOC vs. EBR, $\chi^2(1, N = 151) = 5.55$, $p = .018$, $\Phi = .452$) (Fig 5).

Temperament groups and externalizing symptoms

Although our work was framed in the context of links between temperament and internalizing symptoms, we thought it important to see if a similar association might be observed between temperament and externalizing symptoms. This analysis revealed different effects of temperament group on externalizing symptoms (Fig 4C). Our analyses (ANOVAs) revealed main effects of temperament group on 5-year CBCL Externalizing Problems scores at each age temperament was assessed: T1 $F(2, 338) = 5.65$, $p = .0039$, $\eta^2 = .032$; T2 $F(2, 279) = 13.72$, $p < .001$, $\eta^2 = .090$; T3 $F(2, 271) = 25.21$, $p < .001$, $\eta^2 = .032$. Post-hoc comparisons showed that children in the EBD group showed much higher externalizing scores than children in either the EBR or the IOC group (Fig 4C). The analysis of the effects of temperament group on externalizing disorder diagnosis (ADHD, ODD, and/or CD) revealed convergent results: More children in the EBD group (11.8%) at T2 had an externalizing diagnosis than children in the IOC group (2%) at T2 ($\chi^2(1, N = 134) = 3.90$, $p = .048$, $\Phi = .337$), and more children in the EBD group (12.1%) at T3 had an externalizing diagnosis than children in the EBR group (3.3%) at T3 ($\chi^2(1, N = 181) = 4.86$, $p = .028$, $\Phi = .361$) (SI Fig 3). No diagnostic differences were found among the temperament groups identified at T1.

Trajectory of temperament groups and psychopathology

The next question examined was whether the trajectory of temperament groups matters, e.g., whether children who stayed in a certain group would be more (or less) likely to have a diagnosis of anxiety. We compared the anxiety diagnosis rate and CBCL internalizing scores among 5 groups of children – those who stayed in the same cluster from infancy to three years of

age (i.e., “EBR continued”, “EBD continued”, “IOC continued”), as well as children who were in the EBR or EBD group in infancy or at T2 but transitioned to a different group at T3 (i.e., “EBR transitioned”, “EBD transitioned”). We acknowledge that additional groups are theoretically possible, but the number of participants that make up other groups is insufficient to render useful information.

This analysis showed that the proportion of children having a diagnosis at T4 and the total N of each group were the following: EBR continued (diagnosis rate: 0%, total N = 28), EBD continued (19.4%, 36), IOC continued (0%, N = 5), EBR transitioned (10%, N = 90), EBD transitioned (14%, N = 93) (SI Fig 4A). The EBR continued group had the lowest proportion of anxiety diagnosis, while the EBD group had the highest proportion of anxiety diagnosis. The comparison of the CBCL internalizing scores across groups revealed that children in the EBD continued group had the highest internalizing scores, children in the EBD transitioned group had the second highest internalizing scores, and children in the EBR continued group had the lowest scores (SI Fig 4B).

[Figure 5 goes here]

Relation between missingness in outcome variables and other factors

Approximately half of the children with data at T1 did not participate at T3 or T5 or had missing behavioral data or clinical outcomes, partially due to our original plan to follow about 400 children in later follow-ups. Missing data analysis was conducted to determine whether missingness in the behavioral tasks and clinical outcomes was related to temperament groups or socioeconomic status (SES). The analysis revealed that the distribution of temperament groups and SES were not different between children with and without behavioral or outcome data, *p*s >

.05, indicating that the estimates reported in the previous sections are representative of the cohort. Detailed results for this missing data analysis are reported in SI.

Discussion

In the current study, we used a data-driven clustering algorithm to detect groups of children according to their similarity on ratings of temperament traits in early childhood. Taking advantage of a large longitudinal cohort, we analyzed the stability of temperament groups from infancy to 3 years of age and their predictive association with psychopathology assessed at age 5 years. Three distinct temperament groups were identified across ages: one group (EBR) was characterized by low ratings on negative affectivity and high ratings on surgency/extraversion and orienting/regulation – effortful control; a second group (EBD) showed similarly high ratings on surgency/extraversion but opposite patterns of high negative affectivity and low orienting/regulation – effortful control; and a third group (IOC), which was characterized by low ratings on surgency/extraversion and moderate to high ratings on negative affectivity and orienting/regulation – effortful control. External validation tests revealed that children in these temperament groups showed different behavioral performances in attention disengagement, prosocial, and BI tasks. Although the traits of these groups were stable across ages at the group-level and between boys and girls, individual group membership showed moderate change from infancy to later years but relative stability from ages 2 to 3 years. Membership stability varied by temperament group, with children in the IOC group more likely to transition to other groups, particularly between infancy and later ages. Temperament group was also associated with differences in internalizing symptoms and likelihood of meeting criteria for an anxiety disorder at age 5 years. Specifically, children in the EBD and IOC groups at ages 2 or 3 years were more likely than children in the EBR group to have an anxiety diagnosis at 5 years. Further, at age 5

years, children who were in the EBD group showed higher levels of internalizing symptoms generally, whereas children in either the IOC or EBD group showed higher anxiety symptoms specifically compared to children in the EBR group.

Our finding of three distinct temperament groups provides converging evidence for the existence of different temperament profiles from early childhood. A handful of studies have taken a person-centered approach to study temperament profiles in children. The current findings are consistent with prior studies that observed three to four temperament profiles among infants and preschoolers, even though these studies employed different clustering methods, namely k-means clustering and LPA (Beekman et al., 2015; Komsi et al., 2006; van den Akker et al., 2010). The EBR group identified in the current study shares characteristics with one frequently identified profile that has been labeled in other studies as “positive affect/reactive,” “well regulated,” or “resilient” (Beekman et al., 2015; Komsi et al., 2006; Lin et al., 2018; Prokasky et al., 2017; Scott et al., 2016); the ERD group is comparable to the “negative affect,” “dysregulated,” and “under-controlled” profiles identified in these studies; and the IOC group resembles a previously identified “fearful” (Beekman et al., 2015; Prokasky et al., 2017) or “over-controlled” (Komsi et al., 2006) profile. In contrast, the “high reactive” or “active reactive” profile described previously (Beekman et al., 2015) was not apparent in the current study, possibly due to methodological differences among studies, e.g., LPA vs. graph theory-based community detection. Although the present temperament groups were derived from a community sample, the results were very similar to studies with high-risk samples (Karalunas et al., 2014; Lin et al., 2018) and children of different racial backgrounds (Lin et al., 2021). For instance, Lin and colleagues (2018) identified four temperament profiles in a sample of children with prenatal substance exposure, and three of the four profiles ("moderately low reactive,”

"negative reactive, dysregulated," "high positive affect, well-regulated") had characteristics similar to our IOC, EBD and EBR groups, respectively.

The results regarding the developmental trajectory of these temperament groups suggest that the characterization of temperament profiles is stable in early childhood at the group level; however, profile membership at the individual level changes across ages, becoming more stable with age. While a bit more than 40% of the children stayed in the same group from infancy to age 2 years, more than 60% remained in the same group from 2 to 3 years of age. This finding is consistent with the report by (Beekman et al., 2015), such that the same four temperament profiles were found from 9 to 27 months of age, but profile membership changed drastically across ages. Our results also suggest that stability of membership varies by temperament group. Approximately 70% of infants in the IOC group transitioned to the EBR or EBD group by age 2 years, whereas less than 50% of infants in the EBR or EBD groups transitioned to a different group. Moreover, more than 60% of children in the EBR group and 80% in the EBD group remained in the same community from ages 2 to 3 years (Fig 2). It is plausible that children in the IOC group are more malleable to environmental influences. The IOC children share certain characteristics with the other two groups, such as high ratings on negative affectivity as the EBD group and high ratings on effortful control as the EBR group. This overlap could make them more likely to transition to the other groups. The analysis of the relation between the trajectory of temperament groups over time and psychopathology suggests that the trajectory matters and is related to clinical outcomes, such that remaining in the EBD group across the first three years of life is a risk indicator for the development of internalizing problems, whereas transitioning to a different group by age 3 years suggests reduced risk. In contrast, EBR group stability is protective against developing internalizing problems.

The analyses of laboratory data suggest that our data-driven grouping is associated with underlying differences in cognitive and behavioral functioning, including attention allocation, facial emotion perception, prosociality, and BI. Compared to children in the IOC group, children in the EBR group showed longer dwell time on facial expressions before shifting their visual fixation to the peripheral target. This difference is unlikely due to the EBR children having more developed sustained attention ability but rather suggests their increased interest in looking at social cues (e.g., facial expressions) compared to the IOC children, as no difference was found between their dwell time on the non-face patterns (Fig 3A). Moreover, negative affectivity is likely a dominant factor driving children's looking behaviors in attention-bias tasks using facial expressions. An association between negative affectivity and attention-bias towards threat has been reported (e.g., Nakagawa et al., 2012; Roy et al., 2015). The children in the IOC and EBD groups showed similar levels of negative affectivity, which could explain why they did not show differences in their looking behaviors to facial expressions. Pérez-Edgar and colleagues (2010, 2011) observed that, among children showing attention bias to angry compared to happy faces, BI in early childhood predicted social withdrawal in adolescence; this association was not observed among children showing no bias towards angry faces. Hence, future studies may further examine the moderation effect of attention-bias on the relation between temperament groups and internalizing and externalizing problems.

Prosocial behavior has been positively related to sociability and activity level and negatively related to shyness and negative emotionality in school-age children and adolescents (Carlo et al., 2012; Gross et al., 2015). These prior results support our finding that 3-year-old children in the IOC group, featured by low activity and surgency levels and high negative emotionality, were less likely to help others spontaneously compared to children in the other

groups. However, our results also indicated the IOC children are able to understand others' predicaments but may be too inhibited to offer help without prompting. These relations found between child temperament and prosocial behaviors suggest that researchers may consider child temperament when measuring prosociality, e.g., whether the prosocial tasks are valid or how they should be administered when used with behaviorally and socially inhibited children.

The current BI tasks adopted from Fox et al. (2001) have been used widely in the field. The IOC children's performance in the BI tasks, reflected in high BI composite scores, along with their temperament characteristics, suggest that they resemble the behaviorally inhibited children defined by Kagan, Fox, and colleagues (e.g., Kagan & Snidman, 1999; Fox et al., 2001). This similarity may help explain the instability of the IOC group in early childhood, as BI in toddlerhood is only moderately correlated with BI in later childhood, whereas there appears to be more continuity and stability in BI and its relation to social anxiety by age 5 to 6 years (Fox et al., 2001; Kagan et al., 1984, 1987).

The EBD temperament profile was associated with elevated ratings of general internalizing symptoms and specific anxiety symptoms at age 5 years compared to the EBR profile; moreover, a disproportionate number of children with a diagnosed anxiety disorder at age 5 years were assigned to the EBD group. Notably, the EBD subtype was not synonymous with internalizing problems, as children in the IOC group at age 3 years accounted for approximately 40% of children with an anxiety disorder at age 5 years and demonstrated greater internalizing problems than the EBR group. In addition, the EBD subtype was associated with elevated externalizing symptoms at age 5 years and a higher chance of a lifetime externalizing disorder diagnosis. While the EBD and IOC temperament profiles may both capture core behavioral traits that are early indicators of later internalizing problems, the EBD profile may be

associated with increased risk of both internalizing and externalizing problems, similar to the early signs of childhood irritability (Wakschlag et al., 2018). Such an interpretation aligns with the extant literature suggesting that emotional and behavioral dysregulation in early childhood is associated with irritable and callous phenotypes in later childhood and adolescence (Wiggins et al., 2018; Wakschlag et al., 2018).

In contrast, the IOC profile was specifically related to internalizing symptoms and anxiety disorders. While both the EBD and IOC profiles were characterized by high ratings on negative affectivity, the IOC profile was associated with lower ratings on surgency/extraversion and higher ratings on orienting/regulation – effortful control compared to the EBD profile. These differences suggest that, for IOC children, their ability to control behavior and attention might offer some protection against developing externalizing psychopathology (Eisenberg et al., 2010; Nigg, 2006), whereas their shyness and inhibition might put them at increased risk for developing anxiety problems (Kagan & Snidman, 1999; Pérez-Edgar & Fox, 2005). This interpretation is consistent with the notion that self-regulation and effortful control contribute to preschoolers' social-emotional competence (Rhoades et al., 2009, 2011), and also is in line with the finding that the likelihood of BI children not becoming exuberant and developing externalizing problems is even greater than the likelihood of them developing anxiety problems (Kagan, 2003; Kagan & Fox, 2007). Further, these theories well explain the current finding that the EBR profile, characterized by low negative affectivity and high surgency/extraversion and orienting/regulation – effortful control, was associated with a much lower probability of developing either internalizing or externalizing problems. Overall, our results highlight the importance of considering the co-occurrence of multiple facets of temperament traits in the

prediction of psychopathology rather than focusing on one or two individual dimensions of temperament.

Together, the current findings on differences between temperament groups suggest that, while differences in temperament traits and the heterogeneity in within-group correlations among traits are dimensional (Rothbart et al., 2000), qualitative differences in behaviors do exist between temperament groups (Kagan, 2003), and what makes for distinct temperament groups is how the various dimensional variables “hang together”. For example, children in the EBR and the IOC groups can both score high on effortful control, but they differ on negative affectivity. Similarly, both EBD and IOC groups can be high on negative affectivity, but the IOC group is low on surgency, and thus these two groups of children likely express their negative affect differently – IOC children had lower externalizing problems than EBD children because of their low surgency and high effortful control. These findings speak to limitations in looking at any of these temperament dimensions in isolation because it matters how the child scores on other dimensions to get a more comprehensive picture of the child's "overall temperament." To sum up, it is the combination of scores on dimensional characteristics that can produce more distinct groups, where there can be overlap between groups on individual dimensions.

In the current study, we established an association between temperament profiles in early childhood and later psychopathology. Our results demonstrate that data-driven clustering using a community detection algorithm can be used to predict internalizing (as well as externalizing) problems in children. One major advantage of this approach is that it combines individuals with a given profile of variables (e.g., temperament traits) into one category and focuses on a within-person structure of variables (Bathelt et al., 2018). Thus, using this approach may facilitate the identification of children at risk for developing internalizing as well as other psychological or

neurodevelopmental disorders (Karalunas et al., 2014). Given that the parent-report temperament questionnaires utilized to derive the temperament groups are easy to collect and widely used by research labs and clinics (Rothbart et al., 2001), identifying a child's temperament type may be the first step of screening before more time-consuming behavioral observations and diagnostic evaluations are conducted.

There are a number of limitations to keep in mind when interpreting the current findings. First, our analyses relied on parent-report measures of child temperament rather than laboratory observations. However, the external validation tests indicated that children in these temperament groups showed differences in their behavioral performance in the prosocial, eye-tracking, and BI tasks as anticipated. This is in line with evidence from others demonstrating the predictive and construct validity of the temperament instruments utilized, including prediction of later socioemotional functioning and significant correlations with laboratory-based observations (Gartstein & Marmion, 2008; Rothbart & Bates, 2007). Second, the question of objectivity in measuring child behavior, including temperament, is still open to debate. However, the items on the IBQ-R and ECBQ were designed specifically to reduce the influence of reporter biases by inquiring about concrete child behaviors rather than asking for abstract judgments. Moreover, parent reporting provides the opportunity to gather data on child behavior across various contexts. Importantly, a recent report suggests that maternal characteristics, specifically psychopathology, minimally biases maternal reports of child emotional and behavioral problems (Olino et al., 2021). Prior studies have shown similarities and differences of temperament profiles among racial groups (Chen, Li, & Cao, 2011; Lin et al., 2021). Another limitation of the current study is that most of the children were from White middle to high socioeconomic families. Investigation of relations between temperament profiles and psychopathology in

samples of different cultures and sociodemographic backgrounds would be an important addition to the literature. Finally, researchers who wish to apply the clustering algorithm used in this study should note that the algorithm requires a minimum sample size (roughly 100 – 200) to detect groups and obtain stable results. Hence, data sharing is important and will facilitate researchers and clinicians conducting clustering analysis of child temperament by combining data from individual studies with publicly available datasets. List-wise deletion was applied when analyzing the relations between temperament groups and behavioral performance and clinical outcomes. While list-wise deletion is easy to implement and requires no justification for data imputation, it significantly reduces the number of usable data, and thus alternatives (e.g., path analysis and multiple imputation methods) may be considered when the sample size is limited.

Conclusions

The current study investigated temperament profiles in early childhood, including their ability to predict psychopathology at age 5 years. The likelihood of developing an anxiety disorder and the severity of internalizing symptoms present at age 5 years differed by temperament profile. Our findings highlight the potential utility of this method to facilitate early identification of children at risk for later psychopathology, which in turn can be an initial step of routine screening before more structured evaluation and intervention programs are implemented. Our findings also suggest that prospective studies beginning in infancy that include both psychological and clinical assays are especially important to understand the origins of anxiety and other internalizing problems (Beesdo et al., 2009). Future research may explore the development and maintenance of these temperament profiles throughout childhood and adolescence and examine how profile status may change as a function of environmental factors

(e.g., parenting, culture, socioeconomic status). Such information may inform the development of earlier and more effective risk identification and prevention efforts.

Author acknowledgements

The current work is supported by the National Institute of Mental Health (R01 MH078829; to CAN and MBE) and a Tommy Fuss Center Innovation Award from the Tommy Fuss Center for Neuropsychiatric Disease Research at Boston Children's Hospital (to MBE). We would like to thank all the families for their participation in this study. We would also like to acknowledge the Emotion Project staff past and present for their assistance in data acquisition, preprocessing, and relevant discussion.

References

Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. In *Journal of Child Psychology and Psychiatry and Allied Disciplines* (Vol. 49, Issue 3, pp. 251–275). John Wiley & Sons, Ltd.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles* (Vol. 30). Burlington, VT: University of Vermont, Research center for children, youth, & families.

Ahadi, S. A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the US and China: similarities and differences. *European Journal of Personality*, *7*(5), 359–378. https://doi.org/10.1002/per.2410070506

Bathelt, J., Holmes, J., Astle, D. E., Centre for Attention, L., & Memory, T. (2018). Data-Driven Subtyping of Executive Function-Related Behavioral Problems in Children. *J Am Acad Child Adolesc Psychiatry*, *57*(4), 252-262 e4. https://doi.org/10.1016/j.jaac.2018.01.014

Beekman, C., Neiderhiser, J. M., Buss, K. A., Loken, E., Moore, G. A., Leve, L. D., Ganiban, J. M., Shaw, D. S., & Reiss, D. (2015). The Development of Early Profiles of Temperament: Characterization, Continuity, and Etiology. *Child Dev*, *86*(6), 1794–1811. https://doi.org/10.1111/cdev.12417

Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. *Psychiatric Clinics of North America* , *32*(3), 483–524. https://doi.org/10.1016/j.psc.2009.06.002

Behrendt, H. F., Wade, M., Bayet, L., Nelson, C. A., & Bosquet Enlow, M. (2020). Pathways to

social-emotional functioning in the preschool period: The role of child temperament and maternal anxiety in boys and girls. *Development and Psychopathology*, *32*(3), 961–974. https://doi.org/10.1017/S0954579419000853

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, *9*(2), 291–319. https://doi.org/10.1017/s095457949700206x

Carlo, G., Crockett, L. J., Wolff, J. M., & Beal, S. J. (2012). The role of emotional reactivity, self-regulation, and puberty in adolescents' prosocial behaviors. *Social Development*, *21*(4), 667-685.

Chen, X., Wang, L., & Cao, R. (2011). Shyness-sensitivity and unsociability in rural Chinese children: Relations with social, school, and psychological adjustment. *Child development*, *82*(5), 1531-1543.

Chronis-Tuscano, A., Degnan, K. A., Pine, D. S., Perez-Edgar, K., Henderson, H. A., Diaz, Y., ... & Fox, N. A. (2009). Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(9), 928-935.

Degnan, K. A., Almas, A. N., & Fox, N. A. (2010). Temperament and the environment in the etiology of childhood anxiety. *Journal of Child Psychology and Psychiatry*, *51*(4), 497-517.

Dougherty, L. R., Tolep, M. R., Bufferd, S. J., Olino, T. M., Dyson, M., Traditi, J., Rose, S., Carlson, G. A., & Klein, D. N. (2013). Preschool Anxiety Disorders: Comprehensive Assessment of Clinical, Demographic, Temperamental, Familial, and Life Stress

Correlates. *Journal of Clinical Child and Adolescent Psychology*, *42*(5), 577–589. https://doi.org/10.1080/15374416.2012.759225

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. In *Annual Review of Clinical Psychology* (Vol. 6, pp. 495–525). Annual Reviews. https://doi.org/10.1146/annurev.clinpsy.121208.131208

Fu, X., & Pérez-Edgar, K. (2015). Theories of temperament development. In J. D. Wright (Ed.), International encyclopedia of social & behavioral sciences (2nd ed., pp. 191–198). Oxford: Elsevier.

Gartstein, M. A., & Marmion, J. (2008). Fear and positive affectivity in infancy: Convergence/discrepancy between parent-report and laboratory-based indicators. *Infant Behavior and Development*, *31*(2), 227–238.

Gartstein, M. A., Prokasky, A., Bell, M. A., Calkins, S., Bridgett, D. J., Braungart-Rieker, J., Leerkes, E., Cheatham, C. L., Eiden, R. D., Mize, K. D., Jones, N. A., Mireault, G., & Seamon, E. (2017). Latent profile and cluster analysis of infant temperament: Comparisons across person-centered approaches. *Dev Psychol*, *53*(10), 1811–1825.

Gartstein, M. A., Putnam, S. P., & Rothbart, M. K. (2012). Etiology of preschool behavior problems: Contributions of temperament attributes in early childhood. *Infant Mental Health Journal*, *33*(2), 197–211. https://doi.org/10.1002/imhj.21312

Gross, R. L., Drummond, J., Satlof-Bedrick, E., Waugh, W. E., Svetlova, M., & Brownell, C. A. (2015). Individual differences in toddlers' social understanding and prosocial behavior: disposition or socialization?. *Frontiers in Psychology*, *6*, 600.

Kagan, J., Reznick, J. S., Clarke, C., Snidman, N., & Garcia-Coll, C. (1984). Behavioral inhibition to the unfamiliar. *Child Development*, 2212–2225.

Kagan, J., & Snidman, N. (1999). Early childhood predictors of adult anxiety disorders. *Biological Psychiatry*, *46*(11), 1536–1541. https://doi.org/10.1016/S0006-3223(99)00137-7

Kamphaus, R. W., & Mays, K. L. (2011). Assessment of internalizing behavioral deficits. In *The Oxford handbook of school psychology* (p. 312). Oxford University Press.

Karalunas, S. L., Fair, D., Musser, E. D., Aykes, K., Iyer, S. P., & Nigg, J. T. (2014). Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: Toward biologically based nosologic criteria. *JAMA Psychiatry*, *71*(9), 1015–1024. https://doi.org/10.1001/jamapsychiatry.2014.763

Komsi, N., Raikkonen, K., Pesonen, A. K., Heinonen, K., Keskivaara, P., Jarvenpaa, A. L., & Strandberg, T. E. (2006). Continuity of temperament from infancy to middle childhood. *Infant Behav Dev*, *29*(4), 494–508. https://doi.org/10.1016/j.infbeh.2006.05.002

Laible, D., Carlo, G., Murphy, T., Augustine, M., & Roesch, S. (2014). Predicting Children's Prosocial and Co-operative Behavior from Their Temperamental Profiles: A Person-centered Approach. *Social Development*, 734–752. https://doi.org/10.1111/sode.12072

Lin, B., Lemery-Chalfant, K., Beekman, C., Crnic, K. A., Gonzales, N. A., & Luecken, L. J. (2021). Infant Temperament Profiles, Cultural Orientation, and Toddler Behavioral and Physiological Regulation in Mexican-American Families. *Child Development*.

Lin, B., Ostlund, B. D., Conradt, E., Lagasse, L. L., & Lester, B. M. (2018). Testing the programming of temperament and psychopathology in two independent samples of children with prenatal substance exposure. *Development and psychopathology*, *30*(3), 1023-1040.

Nakagawa, A., & Sukigara, M. (2012). Difficulty in disengaging from threat and temperamental

negative affectivity in early life: a longitudinal study of infants aged 12-36 months. *Behav Brain Funct*, *8*, 40. https://doi.org/10.1186/1744-9081-8-40

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Nigg, J. T. (2006). Temperament and developmental psychopathology. In *Journal of Child Psychology and Psychiatry and Allied Disciplines* (Vol. 47, Issues 3–4, pp. 395–422). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1469-7610.2006.01612.x

Olino, T. M., Michelini, G., Mennies, R. J., Kotov, R., & Klein, D. N. (2021). Does maternal psychopathology bias reports of offspring symptoms? A study using moderated non-linear factor analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/jcpp.13394

Ostlund B, Myruski S, Buss K, Pérez-Edgar KE (2021a). The centrality of temperament to the research domain criteria (RDoC): The earliest building blocks of psychopathology. Development and Psychopathology 1–15.

Ostlund, B. D., Pérez-Edgar, K. E., Shisler, S., Terrell, S., Godleski, S., Schuetze, P., & Eiden, R. D. (2021b). Prenatal substance exposure and maternal hostility from pregnancy to toddlerhood: Associations with temperament profiles at 16 months of age. *Development and Psychopathology*, 1-18.

Perez-Edgar, K., & Fox, N. A. (2005). Temperament and anxiety disorders. *Child Adolesc Psychiatr Clin N Am*, *14*(4), 681–706, viii. https://doi.org/10.1016/j.chc.2005.05.008

Prokasky, A., Rudasill, K., Molfese, V. J., Putnam, S., Gartstein, M., & Rothbart, M. (2017). Identifying child temperament types using cluster analysis in three samples. *Journal of Research in Personality*, *67*, 190-201.

Putnam, S. P., & Stifter, C. A. (2005). Behavioral approach–inhibition in toddlers: Prediction from infancy, positive and negative affective components, and relations with behavior problems. *Child Development*, *76*(1), 212-226.

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: the early childhood behavior questionnaire. *Infant Behav Dev*, *29*(3), 386–401. https://doi.org/10.1016/j.infbeh.2006.01.004

Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., & Leerkes, E. (2014). Development and assessment of short and very short forms of the infant behavior questionnaire-revised. *Journal of Personality Assessment*, *96*(4), 445–458.

Rhoades, B. L., Greenberg, M. T., & Domitrovich, C. E. (2009). The contribution of inhibitory control to preschoolers' social-emotional competence. *Journal of Applied Developmental Psychology*, *30*(3), 310–320. https://doi.org/10.1016/j.appdev.2008.12.012

Rhoades, B. L., Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2011). Examining the link between preschool social-emotional competence and first grade academic achievement: The role of attention skills. *Early Childhood Research Quarterly*, *26*(2), 182–191. https://doi.org/10.1016/j.ecresq.2010.07.003

Rothbart, M. K., & Bates, J. E. (2007). Temperament. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0303

Rubinov, M., & Sporns, O. (2011). Weight-conserving characterization of complex functional

brain networks. *NeuroImage*, *56*(4), 2068–2079. https://doi.org/10.1016/j.neuroimage.2011.03.069

Scheeringa, M. S., & Haslett, N. (2010). The reliability and criterion validity of the Diagnostic Infant and Preschool Assessment: A new diagnostic instrument for young children. *Child Psychiatry and Human Development*, *41*(3), 299–312.

Scott, B. G., Lemery-Chalfant, K., Clifford, S., Tein, J. Y., Stoll, R., & Goldsmith, H. H. (2016). A twin factor mixture modeling approach to childhood temperament: Differential heritability. *Child development*, *87*(6), 1940-1955.

Stifter, C., & Dollar, J. (2016). Temperament and Developmental Psychopathology. In *Developmental Psychopathology* (pp. 1–62). John Wiley & Sons, Inc.

Wakschlag, L. S., Perlman, S. B., Blair, R. J., Leibenluft, E., Briggs-Gowan, M. J., & Pine, D. S. (2018). The neurodevelopmental basis of early childhood disruptive behavior: Irritable and callous phenotypes as exemplars. *American journal of psychiatry*, *175*(2), 114-130.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303. https://doi.org/10.1126/science.1121448

Wiggins, J. L., Briggs-Gowan, M. J., Estabrook, R., Brotman, M. A., Pine, D. S., Leibenluft, E., & Wakschlag, L. S. (2018). Identifying clinically significant irritability in early childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(3), 191-199.

Xie, W., Leppänen, J. M., Kane-Grade, F. E., & Nelson, C. A. (2021). Converging neural and behavioral evidence for a rapid, generalized response to threat-related facial expressions in 3-year-old children. *NeuroImage*, *229*.

Figure Captions

Fig 1. Profiles of ratings on the IBQ-R and EBCQ subscales in the three temperament groups by age and sex as indicated by the community detection algorithm. The solid lines represent the mean z-score of each subscale, and the error bars represent +/- 1 standard error. The subscales are grouped into three different domains, i.e., surgency, negative affectivity (NA), and orienting/regulation (O/R) or effortful control (EC).

Fig 2. Transitions between groups in infancy, 2 years, and 3 years. The thickness of the lines indicates the proportion of children from each group that transitioned to the linked group at an older age. The circular plots show the temperament profile of each group. The innermost circles colored in gray mark -1.5 standard deviations (SD) from the median; the outmost gray circles mark +1.5 SD. The middle gray circles indicate the median. Silhouette plots, shown in the bottom panel, indicate the quality of the clustering solution at each age. Positive values indicate a higher silhouette coefficient, i.e. greater similarity among data in the same group. The dotted vertical line shows the mean value across all data points. Abbreviations: App - IBQ-R Approach scale, Voc - IBQ-R Vocal Reactivity scale, HiP - IBQ-R High Pleasure scale, Smi - IBQ-R Smiling scale, Act - IBQ-R Activity scale, Sen - IBQ-R Perceptual Sensitivity scale, Sad - IBQ-R Sadness scale, Dis - IBQ-R Distress scale, Fea - IBQ-R Fear scale, Rea - IBQ-R Fall Reactivity scale, LoP - IBQ-R Low Pleasure scale, Cud - IBQ-R Cuddliness scale, Ori - IBQ-R Orienting scale, Soo - IBQ-R Soothability scale; Imp - ECBQ Impulsive scale, Act - ECBQ Activity scale, HiP - ECBQ High Pleasure scale, Soc - ECBQ Sociability scale, Ant - ECBQ Positive Anticipation scale, Dis - ECBQ Discomfort scale, Fea - ECBQ Fear scale, Mot - ECBQ Motor scale, Sad - ECBQ Sadness scale, Sen - ECBQ Perceptual Sensitivity scale, Shy - ECBQ Shyness scale, Soo - ECBQ Soothability scale, Fru - ECBQ Frustration scale, Inh - ECBQ Inhibitory Control scale, Att - ECBQ Attention Shifting scale, LoP - ECBQ Low Pleasure scale, Cud - ECBQ Cuddliness scale, Foc - ECBQ Attentional Focus scale.

Fig 3. External validation of the temperament groups identified by clustering analysis. **A.** Children's dwell time (DT) at age 5 years by temperament groups and stimulus type. Children in the EBR group showed greater DT on the stimuli (primarily the faces) compared to children in the IOC group. **B.** Children's prosocial behavior in the Bin episode at age 3 years by temperament groups. The percentiles of EBR and EBD children coded as "spontaneous helper" was greater compared to the IOC children. By contrast, children in the IOC group were more likely to be coded as a "non-helper." There was no difference between groups in the likelihood of being coded as a "prompted helper." C. Children's behavioral inhibition (BI) composite score at age 3 years by temperament groups. Children in the IOC group exhibited greater BI score than children in the other two groups. The temperament groups at age 3 years were used in these external validation tests. The error bars in A and C represent standard errors of mean, and $*p < .05$, $**p < .01$, $***p < .001$.

Fig 4. Internalizing (A), anxiety specific (B) and externalizing (C) symptoms by temperament group. The y-axes show the CBCL Internalizing (A) and Anxiety (B) Problems T-scores. The x-axes

display temperament group by age. Error bands represent +/- 1 standard error. *adjusted $p < .05$, **adjusted $p < .01$, ***adjusted $p < .001$.

Fig 5. Anxiety diagnosis by temperament group and age. The area of the pie represents the percentile of children with (black) or without (green) one or more anxiety diagnoses at age 5 years. The proportion of children having an anxiety diagnosis did not differ by temperament group in infancy but did at age 2 years (EBD > EBR) and at age 3 years (EBD, IOC > EBR).