Longitudinal Stability and Change in Time-Frequency Measures from an Oddball Task During Adolescence and Early Adulthood

Supplementary Material

Stephen M. Malone, Jeremy Harper & William G. Iacono

Contents

1 Method

1.1 Participants

Socioeconomic status. Biological fathers no longer in the home were recruited in addition to 72 step-fathers living with the twins. For present purposes, we took the maximum score across the two on measures of occupational status and educational attainment. As would be expected, given the representativeness of the sample, the majority of parents (56% of each sex) had occupations belonging to the middle three Hollingshead codes. This corresponds to skilled manual labor; clerical and sales workers, technicians, and owners of small businesses; and administrative personnel, small independent businesses and "minor" professionals. Approximately 60% of parents had a high school degree or general equivalency diploma, while approximately 30% of fathers and 28% of mothers had a college or advanced degree. Occupational status was slightly higher among participating mother than among those who did not participate, but this was not true of fathers (Iacono et al., 1999).

1.2 Cumulative drinking assessment

At each assessment, participants were asked about their drinking habits. At the initial assessments, this took the form of a computerized substance use inventory (CSU; Han et al. (1999)). Beginning with the age-17 assessment, participants were interviewed by means of the Composite International Diagnostic Inventory of the Substance Abuse Module (CIDI-SAM; Robins et al. (1990)), a semi-structured clinical interview. We created ordinal measures of four different but related aspects of drinking: the typical nunmber of drinks per occasion; the frequency of drinking; the maximum number of drinks in a 24-hr period; and the number of times drinking to the point of becoming intoxicated. These are described in **Table S1**.

Table S1: Indicators of Alcohol Use

Scale Score	Drinking Frequency	Typical Quantity	Maximum Quantity	Intoxications
0	Never or not in the past year	0	0	0
1	Less than once a month	1-3	1-3	1-5
2	1-3 times per month	4-6	4-6	6-10
3	1-4 times per week	7-10	7-10	11-20
4	Every day or nearly every day	11-20	11-20	21-50
5	2 or more times a day	21-29	21-29	51-149
6	NA	30+	30+	150+

Note: Typical Quantity refers to the number of alcoholic drinks typically consumed when a participant drank. Maximum Quantity refers to the maximum number of drinks consumed in a single 24-hr period. Intoxications refers to the number of times drinking to the point of becoming intoxicated in one's lifetime. The other three measures ask about the participant's lifetime at the age-17 assessment, the first time they were assessed by means of the CIDI-SAM and since the previous assessment thereafter.

1.2.1 Reliability of drink index and twin similarity therein

We computed Cronbach's alpha among the four indicators as a measure of internal consistency. Values ranged from 0.81 to 0.94 (Mdn = 0.895). In **Figure S1** we plot the distribution of scores on this measure for each assessment wave, beginning with the age-14 assessment.

We computed twin ICCs characterizing the degree of similarity of twins at each assessment using the `psych` package. These ranged from 0.52 to 0.68, indicating high levels of twin similarity at each assessment wave.

1.3 PCA and cross-validation

For each assessment wave, we determined an appropriate number of components to retain by means of cross-validation. This consisted of randomly splitting the sample at each assessment wave into two independent sets of twins (or individuals without a matching twin) and splitting the odd- and even-numbered time-frequency bins into sets, thus dividing the 2-dimensional matrix of subjects by time-frequency bins into two

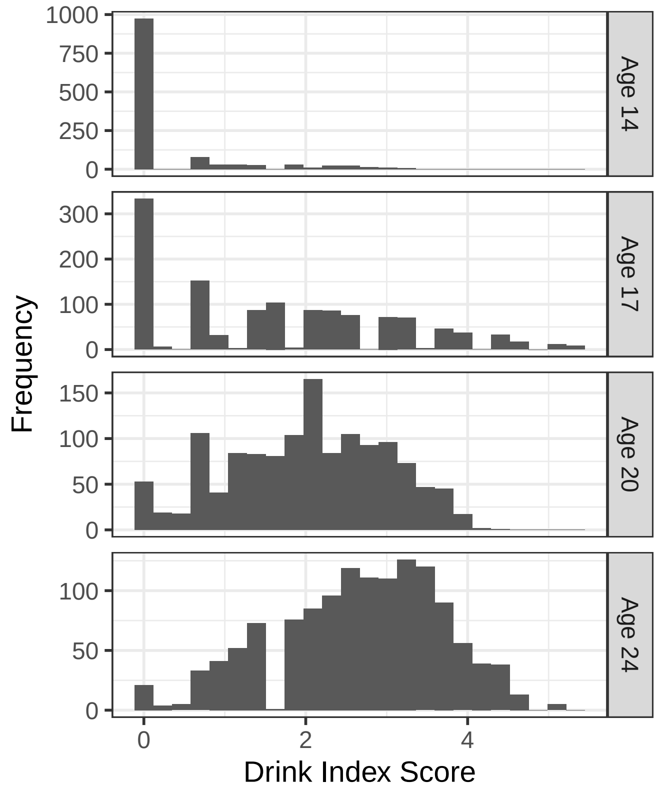

Figure S1: Distribution of drinking measure across assessment waves.

pairs of two mutually exclusive submatrices in a form of bi-cross-validation (Owen & Perry, 2009). For each pair of submatrices, we conducted PCA on one fold and applied the k vectors of weights (right-hand singular vectors) to the other fold, with the number of components retained, k, varying from 1 to 50, and we calculated mean-square error (MSE) of the resulting approximation of the data in the test fold. We repeated this procedure three times, with twin pairs allocated to the two folds using different random number seeds, and averaged MSE across folds and sets. Plots of MSE for each wave resembled scree plots. We therefore adopted the procedure of Helwig and Snodgrass (2019), examining successive differences in the reduction in MSE resulting from retaining $j+1$ components relative to j and determining the value of j when this difference dropped below a proportion of the standard deviation in cross-validated MSE ($\delta * \sigma_{MSE}$), where a value of 0.2 was used for δ (Helwig & Snodgress, 2019).

1.4 Parameterization used in piecewise linear regression models

In the first piece of the model component scores were parameterized as a function, g(h), of age as follows:

$$g(age_{ij}) = \begin{cases} age_{ij} - 11, & if age_{ij} \leq \theta_k \\ 0, & if age_{ij} > \theta_k \end{cases}$$

In the second piece, component scores were parameterized as a function, h(), of age as follows:

$$h(age_{ij}) = \begin{cases} 0, & if age_{ij} < \theta_k \\ age_{ij} - \theta_k, & if age_{ij} > \theta_k \end{cases}$$

In these equations, i is the individual subject, j the assessment wave and θ_k the inflection point for component k estimated in step one. The model intercept in this parameterization represent the expected component score at the inflection point.

As indicated in the manuscript, an alternate parameterization is possible, which uses the same parameterization for the second piece. However, in the first piece, component scores are assumed to be a simple function of centered age: $g(age_{ij}) = age_{ij} - 11$. Thus, the model estimates component score as a linear function of age (centered at age 11) and a second linear function representing the change in this linear function (slope) after the change point. We approximate this parameterization in the manuscript in order to test whether this change in slope was significant, which would support the validity of the piecewise linear model.

2 Results

2.1 Time-frequency EEG activity across assessment waves

Figure 4.1 in the manuscript displays grand mean ERPs for each of the five assessment waves, from age 11 to age 24. The top panel of **Figure S2** recapitulates that figure, with the amplitude of the ERP decreasingly monotonically with age. The rest of the figure displays the mean time-frequency power values from each wave. Several things are noteworthy about these plots. Most of the (scaled) power in these data occurs at low frequencies. This is consistent with previous research on the P3 at posterior electrode sites (e.g., Karakaş et al. (2000)). This low-frequency response extends through much of the time interval, especially at the earliest ages. With development, the response becomes increasingly compact in time. It also begins to extend into higher frequencies, such as slow theta, particularly by the age-20 and age-24 assessments.

2.2 Component congruence

It is evident from **Figure 3** in the manuscript that the component loadings were similar across assessment ages. Tucker's (1951) congruence coefficients quantify how similar a pair of component loadings are. Congruence coefficients equal the cosine of the angle between a pair of loadings. As such, values of 0 indicate that the angle between two sets of loadings is orthogonal, reflecting a complete lack of concordance between them, whereas values of 1 indicate that the angle between them is 0, reflecting identical loadings, or perfect concordance. As reported in the manuscript, we computed congruence coefficients between loadings at successive assessment waves. Components were matched by their timing. Coefficients were uniformly large for matched pairs of components (range, 0.91 to 1), indicating a high degree of congruence. As a rule of thumb, values of 0.95 or greater suggest that the two components can be considered equal (Lorenzo-Seva & Berge, 2006). Coefficients approached unity with age. By contrast, off-diagonal elements in each matrix of congruence coefficients, reflecting unmatched components, were small and approached 0: the median coefficient ranged in absolute value from from 0.01 to 0.04. Thus, matched components were virtually identical and unmatched ones almost completely unrelated, a pattern of results constituting strong evidence that the component structure is equivalent from one assessment age to the next.

2.3 Regression models of component-score change

2.3.1 Attrition and missingness

If missingness depends on values of the dependent measures, then the data are MNAR (missing not at random). There were 54 MTFS participants who completed a laboratory assessment at the intake visit but not thereafter. Component scores at the age-11 assessment did not significantly differ between these subjects and those who completed at least two in-person assessments, p-values ≥ 0.168. This analysis is limited by the small number of participants completing only the age-11 assessment. We therefore also assessed component score as a function of the number of missed assessments, treated as an ordinal measure. Results were very similar, with a statistically insignificant association between the number of waves missed and component score, p-values ≥ 0.508. These analyses are certainly not definitive, but they nevertheless provide some evidence that data were MAR and therefore "ignorable."

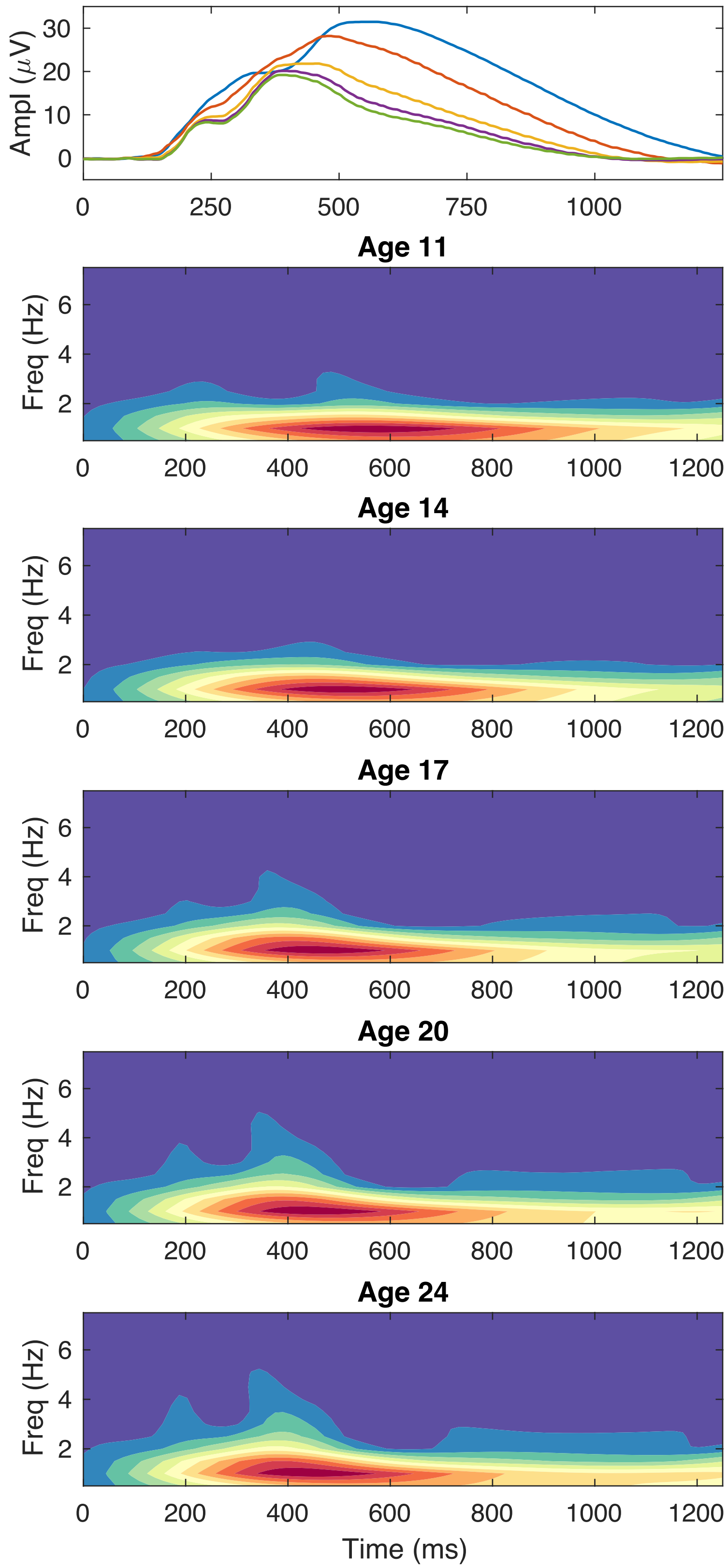

Figure S2: Grand mean ERPs and time-frequency activity across assessments waves.

2.3.2 Comparison between the piecewise model and curvilinear models of change in component scores

Although the fit of the piecewise model to the observed data was compelling, we fit alternate models to component scores for comparison. In two, we treated component score as a quadratic (parabolic) or cubic function of age. We also treated component score as an inverse function of age and as characterized by a (nonlinear) asymptotic function (Luna et al., 2004). In **Table S2**, we present the relative weight of evidence in favor of each model, which are derived from corrected AIC and BIC values. These correspond to the probability of each model, given the data (Anderson, 2008). We recomputed AICc and BIC for the piecewise linear model to account for the change point parameter, which is estimated separately.

Component 2's trajectory was characterized by the smallest amount of change, and BIC weights favored this model slightly over the piecewise linear model, with a probability of 0.64 versus 0.35. Akaike weights indicated the reverse, with a probability of 0.92 for the piecewise linear model and only 0.06 for the quadratic model. For the other five components, model evidence unequivocally favored the piecewise linear model, with probabilities greater than 0.96. These results support our inference that a piecewise linear model is appropriate for these data. **Figure S3** depicts the trajectories predicted by each different curvilinear model as well as the BIC value corresponding to each model, which are overly smooth relative to the observed means.

Table S2: Weight of evidence for different models of change in component scores.

	BIC					AICC				
Component	Piecewise	Quadratic	Cubic	Inverse	Asymptotic	Piecewise	Quadratic	Cubic	Inverse	Asymptotic
PC1	1.	0.	0.	0.	0.	1.	0.	0.	0.	0.
PC2	0.351	0.641	0.009	0.	0.	0.917	0.060	0.023	0.	0.
PC3	1.	0.	0.	0.	0.	1.	0.	0.	0.	0.
PC4	0.999	0.	0.001	0.	0.	0.999	0.	0.001	0.	0.
PC5	1.	0.	0.	0.	0.	1.	0.	0.	0.	0.
PC6	0.966	0.	0.034	0.	0.	0.966	0.	0.034	0.	0.

Note: BIC is the Bayesian Information Criterion (Schwarz, 1978), AICC is the bias-corrected Akaike Information Criterion (Hurvich & Tsai, 1989). Models differ from one another with respect to the function of age used to represent component score. 'Inverse' models component score as an inverse function of age (cf. Luna et al., 2004). 'Asymptotic' is a nonlinear asymptotic regression similar to the three-parameter exponential function used by Luna and colleauges. Fit statistics for the piecewise linear model have been adjusted for the additional parameter implicit in this model, the unknown change point, which we estimated separately. Values reflect the probability of each respective model, given the data (Anderson, 2008).

2.3.3 Sex differences in rates of change: Model evidence

Bayes factors unambiguously favor a model with a main effect of sex only for components 1–4 and suggested a meaningful sex differences in rates of change for component 6 only. In light of the marked sex differences observed by Chorlian and colleagues (2015), we thought it prudent to examine evidence supporting the two models derived from a different information-theoretic fit statistic, the bias-corrected AIC. Although AIC and BIC differ in important ways, from a practical perspective AIC will tend to favor more complex models than BIC. This makes it more sensitive to sex differences in rates of change in component score. **Table S3** presents posterior model probabilities, based on so-called Akaike weights, or the weight of evidence in favor of a given model (Burnham & Anderson, 2004). These clearly favor the sex differences model for components 5 and 6, with a lesser degree of evidence in favor of sex differences in rates of change in scores on components 1 and 2. There may be true differences in component-score trajectory between males and

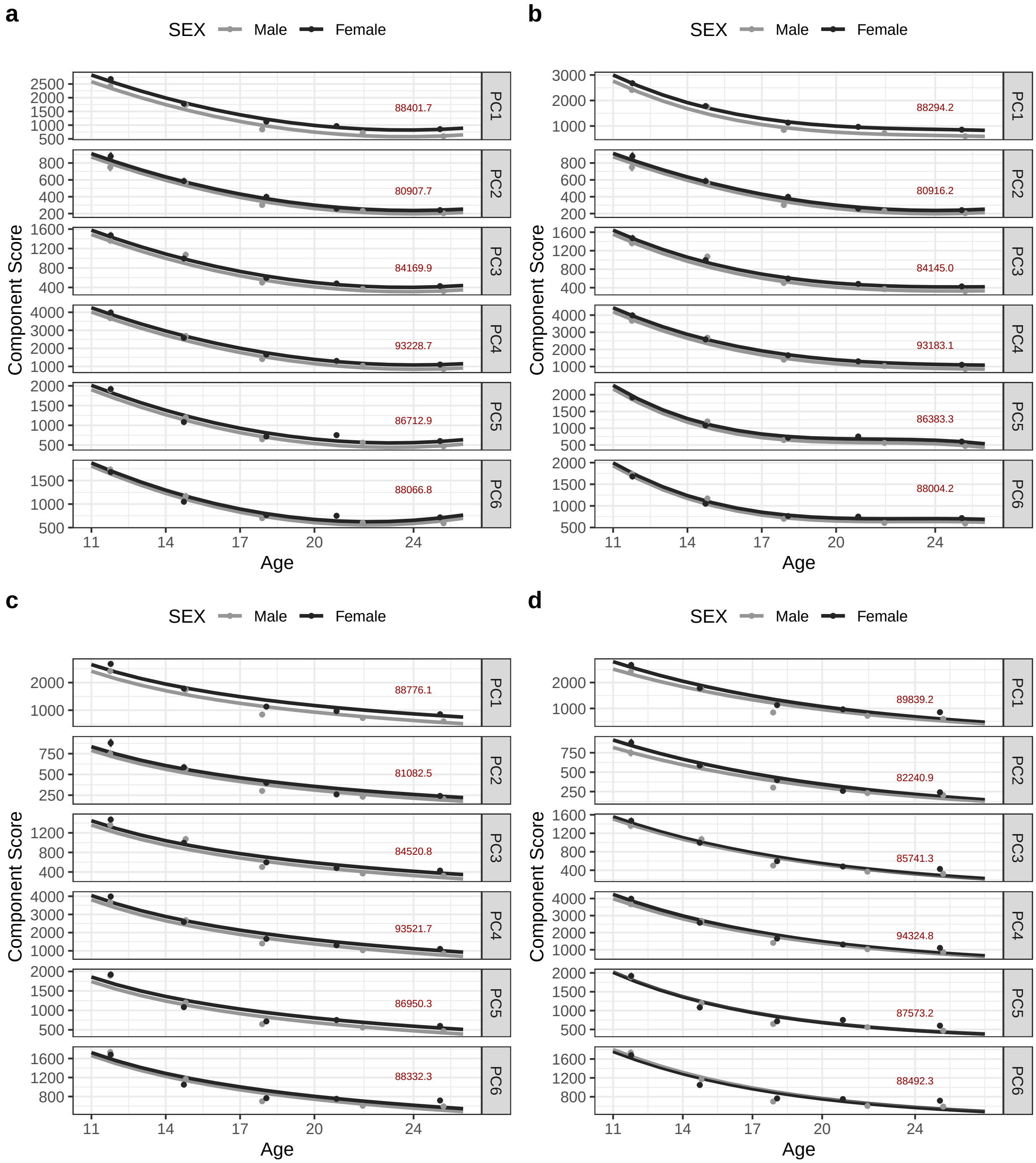

Figure S3: Model-implied trajectories of change in time-frequency components, plotted separately for males and females. Component scores were modeled as second (a) and third (b) order polyonomial functions of age or as an inverse (c) or asymptotic (d) function or age. Observed (measured) means are represented by filled circles, with 95% confidence intervals around them as vertical lines.

Table S3: Log-likelihoods and model evidence for sex differences in rates of chanage in component scores.

	Log-likelihood		AICC Model Probability		BIC Model Probability	
	Baseline	Sex Differences	Baseline	Sex Differences	Baseline	Sex Differences
PC1	-44015.2	-44011.9	0.212	0.788	0.995	0.005
PC2	-40398.2	-40394.3	0.138	0.862	0.992	0.008
PC3	-41971.5	-41970.4	0.715	0.285	0.999	0.001
PC4	-46528.2	-46527.8	0.822	0.178	1.000	0.000
PC5	-43114.0	-43105.4	0.001	0.999	0.515	0.485
PC6	-43942.5	-43925.1	0.000	1.000	0.000	1.000

Note: AICC and BIC Model Probabilities represent the relative weight of evidence in favor of a given model. Weights are normalized to 1, so these can be thought of as probabilities. The Sex Differences model includes an interaction between sex and each slope parameter in the piecewise linear model, and it therefore assessed whether there are sex differences in the rates of (linear) change.

females. In particular, the rate of change in component scores was slightly less among females before the inflection point and slightly greater after it. However, p-values associated with the various slope parameters, unadjusted for the number of tests, were less than .05 for only three out of 12 coefficients. These therefore reflect small sex differences in linear rates of change.

2.3.4 Influence of pubertal status on component scores

The total amount of change during this span is almost identical for the two sexes, differing by no more than 0.2%, but for females the majority of change occurred early in adolescence, between 11 and 14, for all components (as much as 70%), whereas for males, the majority of change in components 1–4, as much as two-thirds, occurred between 14 and 17, and the relative amount of change during this interval was greater among males relative to females for all components. (See component 3 in **Figure S3**, for example. The direction of the male-female difference in mean scores for this component flipped between the age-11 and age-14 assessments and again between age 14 and age 17.) We speculated that this might be due to sex differences in the timing of puberty and rates of maturation. Although not commonly considered in relation to EEG measures, there is nevertheless evidence of effects of pubertal hormones on measures of brain organization and in shaping neural circuits (Bedny et al., 2018; Schulz et al., 2009), as well as sex differences in white matter microstructure (Ho et al., 2020). Testicular hormones during puberty appear particularly important for organizing synaptic plasticity in the hippocampus (Schulz et al., 2009), a possible source of the P3 (Halgren et al. (1980); see also Polich (2007)). Sex differences in the onset and course of pubertal development might result in subtly different trajectories of measures of brain function between male and female adolescents. The Pubertal Development Scale (PDS; Petersen et al. (1988)), a self-report measure of bodily changes related to puberty, had been administered at the first two assessment waves as part of the comprehensive assessment of study participants. We therefore conducted follow-up analyses to determine whether variation in pubertal status might account for sex differences in component scores. We caution that these analyses are *post-hoc* and entirely exploratory.

A small number of participants were missing PDS ratings (n = 69), almost all from the intake assessment. Scores were fixed at the maximum score for the age-17 and subsequent assessments. As described in the manuscript, we obtained robust evidence of sex differences in trajectories of component scores only for component 6, the late slow wave component. After we adjusted for scores on a scale of pubertal development in post-hoc analyses, the sex by age interaction effects were not significant. However, after we adjusted for pubertal status, the sex difference in initial slope became marginally significant for component 2, with a p-value adjusted for the 12 tests of sex by age interactions using Holm's multiple comparison procedure

(Holm, 1979) equal to 0.049. We present the model-predicted trajectories in **Figure S4**. Component score trajectories were predicted by a model with a piecewise function of age and a main effect of pubertal status for all by component 2. For component 2, we used parameters from a model with sex by age interactions. To generate predicted scores, we created integer-valued age bins from 11 to 26 and computed the average PDS score for each age bin, along with a dummy variable coding for sex (female = 1). These data were combined with model parameters to generate predicted scores. Model-implied trajectories for component 2 conform more closely to the observed means at the earliest ages, lending credence to the inference that there are sex differences in scores on this component early in adolescence (**Figure S4**). Thus, adjusting for pubertal status allowed a sex difference to emerge. However, we caution that this finding is only suggestive, given the post-hoc nature of the analysis in the first place.

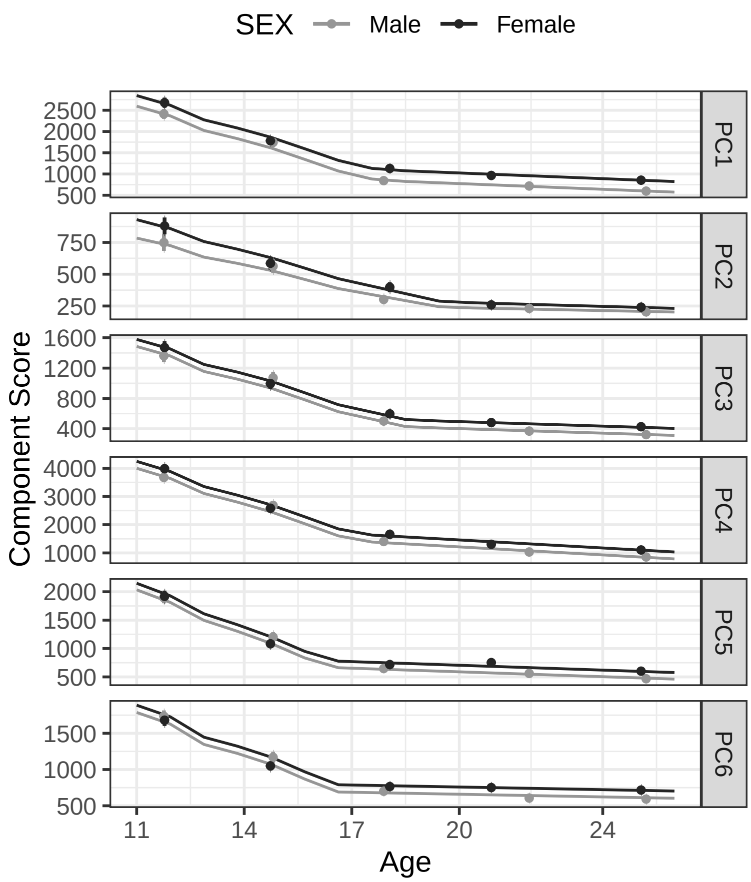

Figure S4: Model-implied trajectories of change in time-frequency components. The change point, or knot, was estimated separately for each component. Trajectories were predicted by a piecewise regression model with a main effect of sex and pubertal status, except component 2, theh trajectory of which was generated by a model including sex by age interaction terms. Observed (measured) means are represented by filled circles, with 95% confidence intervals around them as vertical lines. The scale of the ordinate was allowed to vary across plots to emphasize detail in model-predicted trajectories.

2.4 Developmental change in the context of component stability

Our results indicate that the structure of the time-frequency representation of EEG activity was highly stable from one assessment wave to the next. Nevertheless, significant change is evident, in the grand-mean ERP in particular, but to a lesser degree in the component loadings themselves. **Figure S5** depicts the components organized by component, rather than assessment wave, as in the manuscript, to facilitate visually assessing the degree of stability as well as subtle change in component timing. We conducted two *ad hoc* analyses to further characterize sources of stability and change in these data, which we describe here.

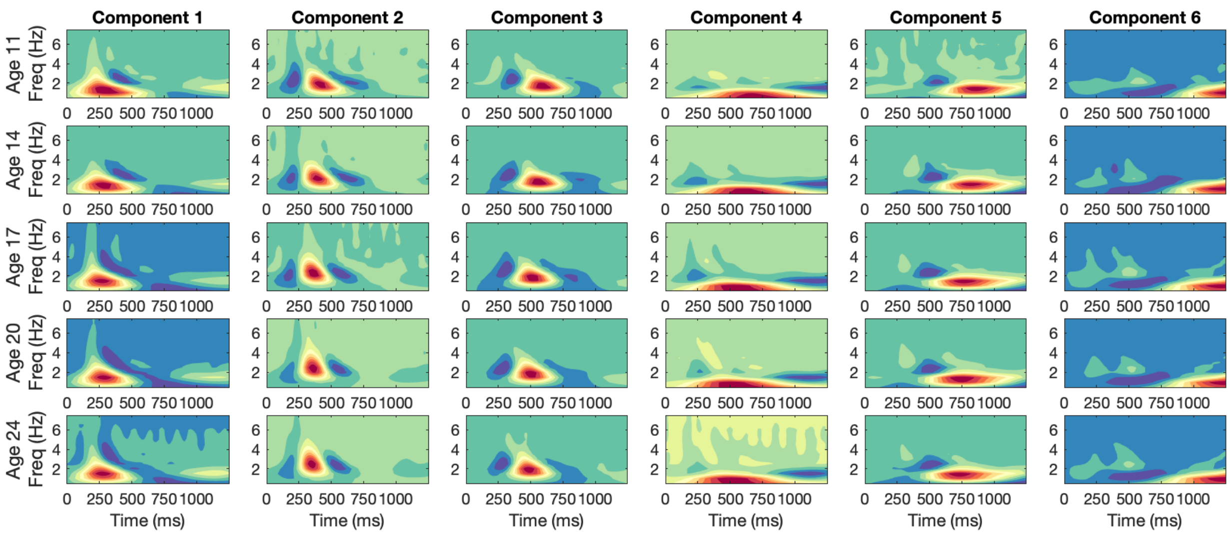

Figure S5: Component loadings across assessment ages. The same heatmaps as in **Figure 3** in the manuscript, but organized by wave, so as to facilitate visually assessing the degree of similarity across assessment waves as well as any change.

2.4.1 Visualizing change in component loadings across age

In the first, we attempted to formalize shifts in timing and frequency that are evident from scanning down the columns of **Figure 3** in the manuscript. We weighted each individual participant's time-frequency energy data matrix elementwise by the corresponding component loading, separately for each component, and determined the bin at which the weighted value was maximal. That is, we multiplied together each element in a row of the the $N \times p$ matrix of time-frequency activity, where each row, n, represents a participant, and the corresponding element in the k^{th} column of the $p \times k$ matrix of component loadings on the p time-frequency bins. We then determined the maximum, akin to determining the latency of an ERP component but in time-frequency space. This approach has the advantage of allowing us to characterize inter-individual variability.

Mean time-frequency peaks, or maxima, are plotted in **Figure S6**. Filled circles represent the mean across participants for a given component, and these are connected by a line, with an arrowhead pointing to the age-24 assessment wave. The size of the circles equals the square root of Wilks's (1932) generalized variance (GV) of their bivariate (time and frequency) distribution. The square root of GV is proportional to the space spanned by points in time-frequency space and is thus a scalar measure of dispersion. The figure indicates systematic shifts toward earlier maxima for all components and toward higher frequencies for the first two components.

2.4.2 Near-perfect component stability emerges from small change over the course of adolescent development

Because these shifts reflect changes in loadings with development, in the second analysis we computed congruence coefficients as a function of the number of assessment waves between loadings being compared. In the manuscript and in **Figure ??**, we report congruence coefficients between successive ages, or assessments only three years apart. Here, we computed congruence coefficients between loadings from assessments six,

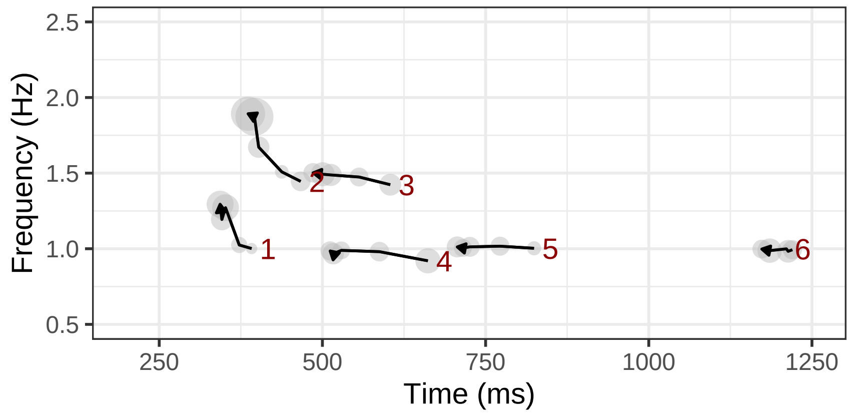

Figure S6: Trajectories of change in component locations (loadings) in time-frequency space. Each circle represents the mean maximum loading for a given component-wave in terms of its time and frequency bin, which have been transformed into time in milliseconds and frequency in Hz. The size of the circle corresponds to the square root of the generalized variance, reflecting the amount of variability, or spread, in component maxima. For each component, its number marks the age-11 value, successive assessment ages are connected by a line and an arrowhead points to the age-24 value.

nine and 13 years apart – i.e., between age 11–age17, age 14–age 20 and age 17–age 24 loadings, between age 11–age 20 and age 14–age 24 loadings, and between age 11–age 24 loadings. We averaged across matrices of coefficients when more than one reflected a given time lag. These average congruence coefficients increased monotonically: With decreases in the time interval between the loadings being compared, the median value increased from 0.89 to 0.92 to 0.95 to 0.98. Thus, the near-perfect congruence in component structure evident in **Figure ??**, across three- to four-year intervals, emerged gradually over the course of the 13-year span of development for those participating in this study.

References

Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence* (Vol. 31). Springer.

Bedny, M., Paus, T., Doesburg, S. M., Giedd, J., Hashemiyoon, R., Kold, B., Purdon, P. L., Rakic, P., & Sisk, C. L. (2018). *Emergent brain dynamics: Prebirth to adolescence* (A. A. Bensaich & U. Ribary, Eds.; pp. 179–193). MIT Press. https://doi.org/10.7551/mitpress/11957.003.0013

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Chorlian, D. B., Rangaswamy, M., Manz, N., Kamarajan, C., Pandey, A. K., Edenberg, H., Kuperman, S., & Porjesz, B. (2015). Gender modulates the development of theta event related oscillations in adolescents and young adults. *Behavioural Brain Research*, *292*, 342–352. https://doi.org/10.1016/j.bbr.2015.06.020

Halgren, E., Squires, N., Wilson, C., Rohrbaugh, J., Babb, T., & Crandall, P. (1980). Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science*, *210*, 803–805. https://doi.org/10.1126/science.7434000

Han, C., McGue, M. K., & Iacono, W. G. (1999). Lifetime tobacco, alcohol and other substance use in adolescent minnesota twins: Univariate and multivariate behavioral genetic analyses. *Addiction*, *94*(7), 981–993. https://doi.org/10.1046/j.1360-0443.1999.9479814.x

Helwig, N. E., & Snodgress, M. A. (2019). Exploring individual and group differences in latent brain networks using cross-validated simultaneous component analysis. *NeuroImage*, *201*, 116019. https://doi.org/10.1016/j.neuroimage.2019.116019

Ho, T. C., Colich, N. L., Sisk, L. M., Oskirko, K., Jo, B., & Gotlib, I. H. (2020). Sex differences in the effects of gonadal hormones on white matter microstructure development in adolescence. *Developmental Cognitive Neuroscience*, *42*, 100773. http://www.sciencedirect.com/science/article/pii/S1878929320300219

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70. https://www.jstor.org/stable/4615733

Iacono, W. G., Carlson, S. R., Taylor, J., Elkins, I. J., & McGue, M. (1999). Behavioral disinhibition and the development of substance-use disorders: Findings from the Minnesota Twin Family Study. *Development and Psychopathology*, *11*(4), 869–900. https://doi.org/10.1017/S0954579499002369

Karakaş, S., Erzengin, O. U., & Başar, E. (2000). A new strategy involving multiple cognitive paradigms demonstrates that ERP components are determined by the superposition of oscillatory responses. *Clinical Neurophysiology*, *111*, 1719–1732. https://doi.org/10.1016/s1388-2457(00)00418-1

Lorenzo-Seva, U., & Berge, J. M. F. ten. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*, 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372. https://doi.org/10.1111/j.1467-8624.2004.00745.x

Owen, A. B., & Perry, P. O. (2009). Bi-cross-validation of the SVD and the nonnegative matrix factorization. *The Annals of Applied Statistics*, *3*. https://doi.org/10.1214/08-aoas227

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence*, *17*, 117–133. https://doi.org/10.1007/bf01537962

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Robins, L., Cottler, L., & Babor, T. (1990). Composite international diagnostic interview-expanded substance abuse module (CIDI-SAM). *St. Louis, MO: Authors.*

Schulz, K. M., Molenda-Figueira, H. A., & Sisk, C. L. (2009). Back to the future: The organizational–activational hypothesis adapted to puberty and adolescence. *Hormones and Behavior*, *55*, 597–604. https://doi.org/10.1016/j.yhbeh.2009.03.010

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies.* Educational Testing Service Princeton NJ.

Wilks, S. S. (1932). *Certain generalizations in the analysis of variance.* *24*, 471. https://doi.org/10.2307/2331979