Response to reviewers

Reviewer #1:

This paper reports analyses from a Monetary Incentive Delay (MID) fMRI task within the IMAGEN cohort at two time points (cohort ages 14 and 19). Both general task-activation and seeded task-based functional connectivity from the ventral striatum are explored as well as inter-individual difference associations with alcohol use (via the AUDIT). This paper has notable strengths including the large longitudinal design, the use of a field-standard neuroimaging task, and well-motivated and interesting neurodevelopmental questions. Enthusiasm is limited however by a number of areas for clarification and concern, including a lack of cohesion integrating the many well-performed, but varying analyses, inconsistency/lack of clarity on certain modeling aspects (e.g., missing data), and room for improvement in figure displays. My more specific comments are listed below.

Primary Comments:

1. This manuscript presents a complementary set of analyses from previous work from this group (Cao et al., 2019) and generally does well to distinguish the current project that includes T1 and T2 time points from this previous work (time point 1 only). However, it would assist the reader to include a bit more clarity on the relationships between these two papers. For example, the sample sizes at T1 appear to differ?

Response: We have clarified this in the Method section (see lines 338-349).

> ## Materials and Methods
>
> ## Brain response and VS connectivity
>
> …which was included as a covariate in the subsequent statistical analysis of brain data.
>
> Compared to the proceeding study that examined the brain response and seed-based functional connectivity during different stages of the MID task (Cao et al., 2019), the current study focused on reward anticipation but additionally examined graph theory metrics of the task-modulated network. Apart from that, there were several improvements in the current study. The previous study used 6 regressors (3 translations 3 rotations) in the individual level models. The current study included 21 movement-related regressors to better account for head motion at the individual level. Furthermore, the mean FD was calculated and included in the group level models to address head motion at the group level analyses in the present study. In addition to the image quality, the current study additionally excluded participants who did not meet the behavioral inclusion criteria and those who failed in task-modulated network construction, which resulted in a smaller number of participants at T1 when compared to that of the previous paper.

2. Overall, there are many potentially complementary analyses performed but the rationale and logic for all analyses is not entirely clear. Some systems-level neuroscience justification for the potentially complementary insights to-be-derived from task-activation, task-connectivity, and graph-based metrics in the introduction and conclusion would be very useful. As a larger, but similar point, this manuscript would broaden its appeal with more clarity on the description of neuroimaging specific terms and analyses.

Response: We have addressed this comment in the Introduction (see lines 139-224). We attempt to explain why seed-based connectivity and whole-brain connectivity measures of the brain can increase our understanding of the brain's reward system and its maturation.

Introduction

… However, it is unknown how the neural correlates underlying reward anticipation in this same sample might change between 14 years old and young adulthood five years later at age 19.

Previous studies on the adolescent reward system emphasized the brain response in isolated regions such as the ventral striatum (VS) (Bjork et al., 2010; Lorenz et al., 2014). In a step towards a larger-scale understanding of the reward system, a seed-based functional connectivity analysis can examine if correlated activity between a seed region and the rest of the brain varies as a function of an experimental manipulation and varies with ageing. The functional coupling of brain responses between areas suggests their involvement in the same underlying functional processes (Lv et al., 2018). Studies have shown that functional connectivity can be found with a region that does not show a significant increase in task-related activity (Cao et al., 2019; Di and Biswal, 2019), indicating that a broader involvement of brain regions can be uncovered by a functional connectivity analysis. For instance, we have previously reported connectivity between bilateral VS and regions that play a role in attention to and integration of salient information (e.g., middle, inferior frontal gyrus, angular gyrus, inferior parietal gyrus, insula and putamen) during reward anticipation in adolescents (Cao et al., 2019). Thus, an examination of VS connectivity can probe the involvement of the non-hedonic components that are linked to the VS activity such as mobilization of attentional or cognitive resources.

Seed-based functional connectivity analysis are, however, restricted to a limited number of seed regions that are chosen by researchers (Stevens, 2016; Cao et al., 2019). As brain function is better characterized as an integrated network (Lv et al., 2018), several graph theory-based metrics (e.g., the network strength, shortest path length, clustering coefficient, efficiency) can be used to assess brain network properties in terms of topographical organization and interregional connectivity. Brain intrinsic functional connectivity undergoes dramatic changes during maturation, with increased integration and segregation facilitating network efficiency (Fair et al., 2007). In line with this finding, adults showed a more flexible and specialized intrinsic functional network compared to adolescents and children (Ernst et al., 2015). Even though the dynamics of functional connectivity can be investigated using resting-state fMRI, the task-modulated functional network, in which high-level task demands are accommodated by context-specific modulations, may offer more insights over the intrinsic functional network (Mennes et al., 2013; Di and Biswal, 2019; Finn, 2021). Therefore, an examination of task-modulated functional network properties can provide insights into the brain's functional integration and topographical organization under the task demand such as reward anticipation. However, no study has

yet applied this type of network analysis in the MID task to characterize the development of the reward system.

Here, participants from a large-scale longitudinal fMRI study (IMAGEN) performed the MID task during …

3. I was a bit confused as to why certain missing data was allowed, where imputation was done, and where complete cases were required. Some strategies for missing data seem very clear. For example, imputation for the years between visit which generally had a small window to begin with. Others are far less clear. For example, why the AUDIT analysis was only complete cases/data at both time points. It is possible [maybe likely], the most clinically acute participants were more likely to drop out of the study from Time 1 to Time. Removal of these subjects likely reduces variance at T1.

Response:

(1) Imputation for the missing data.

The number of years between two visits was the only variable that was imputed for participants. This variable was included as a nuisance regressor in the LME models to capture the individual differences in the length of interval between T1 and T2 and, as noted by the reviewer, there was not a lot of variation in this variable due to the 5-year follow-up design. However, imputation for the AUDIT score could be quite inaccurate as the missing AUDT score at T2 could range from 0-40. As the AUDIT score was used to define groups for the alcohol effects analysis, imputation for AUDIT would not be advisable. Therefore, we focused on the participants who completed AUDIT at both time points rather than imputing it.

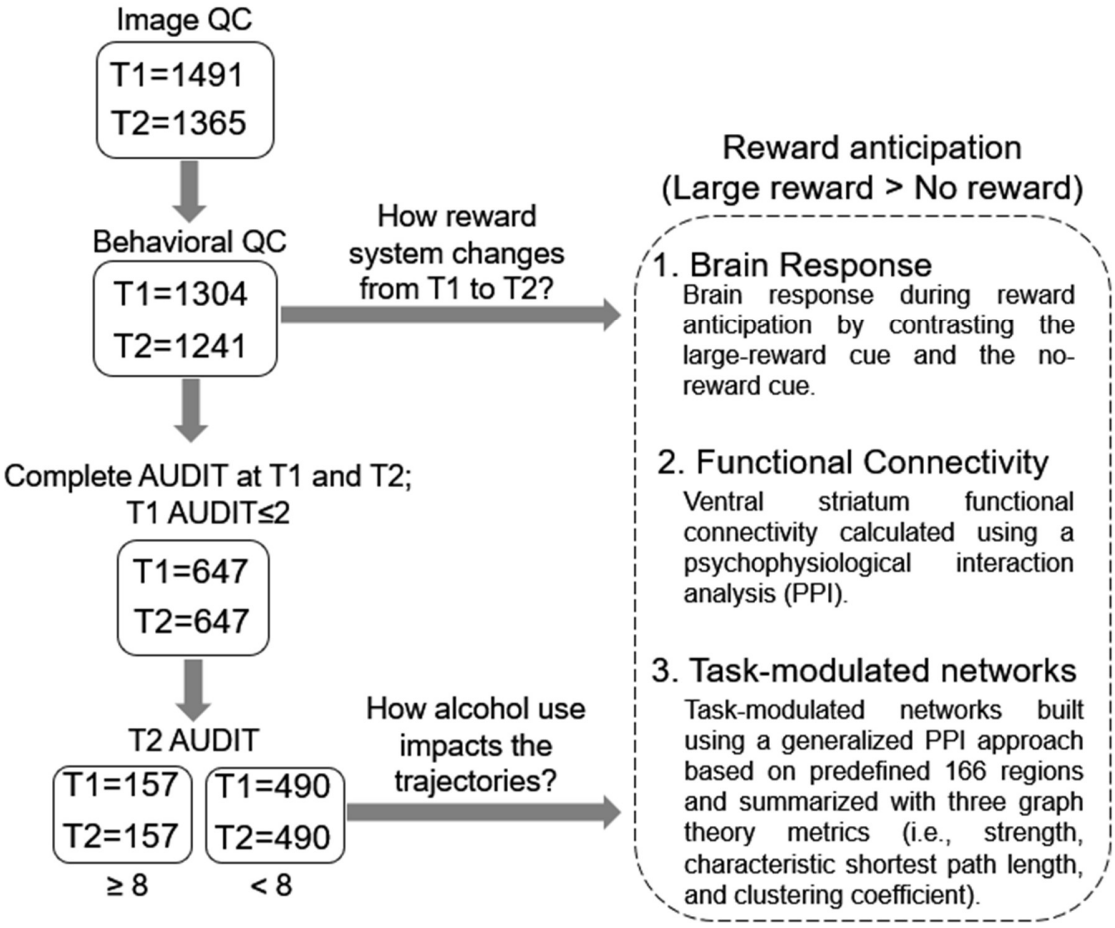
We have clarified these points in the manuscript (see lines 378-396) and additionally included a flow chart that describes the samples included in each analysis.

Materials and Methods

Statistical Analysis

The analytical strategies and the numbers of participants included in each analysis are illustrated in Figure 2. There were 1491 participants at T1 and 1365 participants at T2 who passed image quality control (QC) and 1304 and 1241 participants who further passed behavioural QC and the latter groups were included in the analysis examining the trajectories of brain response, functional connectivity of the VS, and graph theory metrics of the task-modulated network. The number of years between two visits was imputed for participants missing this variable (this was a nuisance covariate of no interest with a tight distribution around five years given the study design). There were 674 participants who met the inclusion criteria based on the completeness and cutoffs of AUDIT, which entered the second set of analyses assessing the interactive effects between alcohol use and brain measures trajectories.

Figure 2. Analytical strategies on brain measures. There were 1491 participants at T1 and 1365 participants at T2 who passed image quality control (QC) and 1304 and 1241 participants who further passed behavioural QC. These participants were included in the analyses examining the trajectories of brain response, functional connectivity of the ventral striatum, and graph theory metrics of the task-modulated network. 674 participants met the inclusion criteria based on the completeness and cutoffs of the Alcohol Use Disorders Identification Test (AUDIT) and were included in the second set of analyses assessing the interactive effects between alcohol use and brain measure trajectories.



(2) Attrited participant at T2.

There are two considerations for setting a low AUDIT at T1: (1) participants who reported a high AUDIT score may have already experienced brain changes correlated with alcohol use; (2) we were interested in comparing participants who increased their drinking to a certain risk level at T2 versus those did not. We agree that it is possible that the most clinically acute participants were more likely to drop out at T2. However, given that our analytic design already restricted our focus to participants with low AUDIT (⩽2) at T1, the exclusion of attrited participants is unlikely to reduce variances at T1. We have performed additional analysis to confirm this (lines 477-480).

Materials and Methods

Alcohol use effects

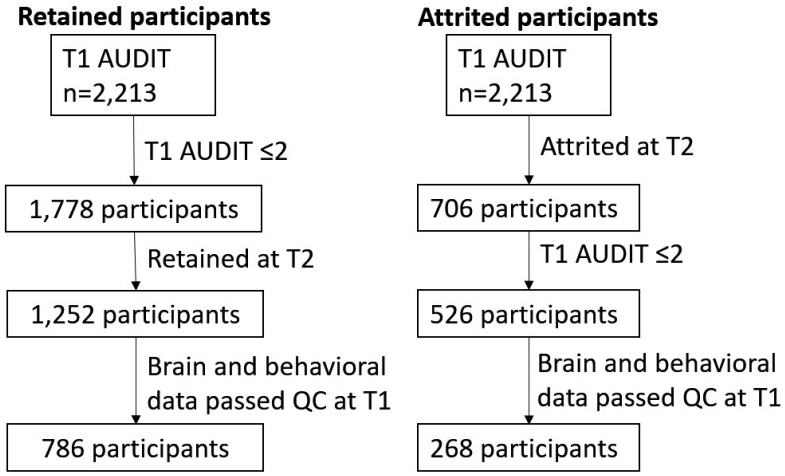
… Those who had a low AUDIT score at T1 (≤2) and T2 (≤7) were designated the low alcohol use group (LA)… Given that our analytic design already restricted our focus to participants to participants with low AUDIT (⩽2) at T1, the exclusion of participants who had dropped out of the study by T2 was unlikely to reduce the variance in AUDIT at T1. Additional analyses were performed to test this in the Supplementary Information.

Supplementary Information

Effects removing attrited participants on variance of T1 data

We tested whether removing the participants who were attrited at T2 significantly changed the variance of data at T1. The strategies to identify the attrited and retained participants are shown in **Figure S11.** We compared the retained group (Group 1, n=786) and retained plus attrited group (Group 2, n=1054) on the variance of brain response, VS functional connectivity and graph theory metrics at T1. Specifically, we tested whether the variance ratio between the two groups was significantly different from 1 (i.e., a two-sided F test). A significant result means removing the attrited participants changed the variance of the brain measure at T1.

Figure S11. Strategies to identify the attrited and retained participants.



As shown in **Figure S12** and **Table S4**, very few voxels survived the FDR-corrected $p<0.05$ in the voxel-wise analysis of the brain response and VS functional connectivity between the retained and retained plus attrited participants. Given the location of these voxels, the impact of removing attrited participants was unlikely to contribute to the existing findings of alcohol effects on right VS connectivity with left middle temporal gyrus (Peak MNI: -57 -61 -2).

Figure S12. A. Histogram of the variance ratio (i.e., F value) between the retained participants and retained plus attrited participants for brain response and functional connectivity of the VS. The histograms suggest the two groups have similar variance for most of the voxels with F values centered at 1. B. Voxels showing a significant difference in the variance between the two groups (i.e., F value is significantly different from 1, FDR-corrected $p<0.05$).

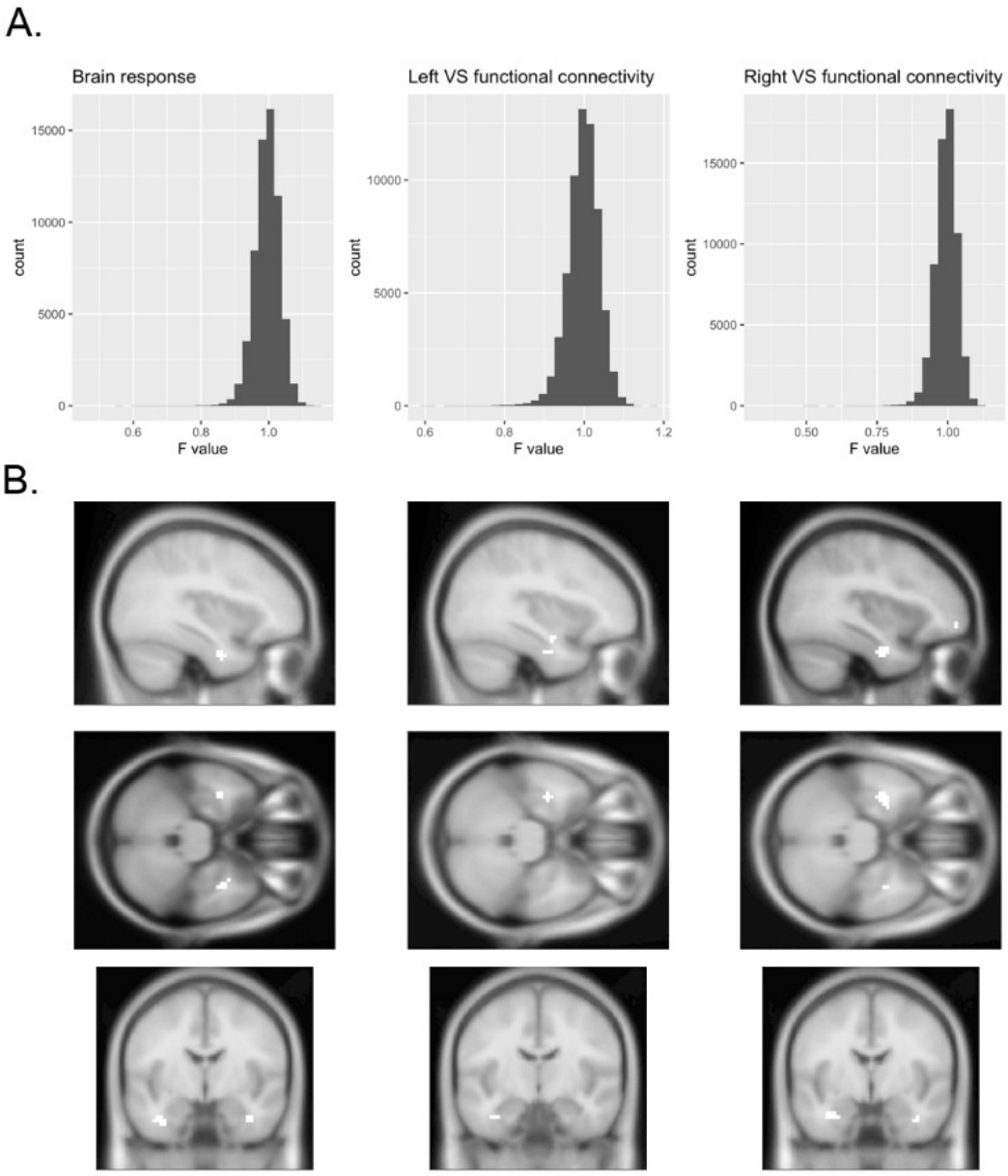


Table S4. Regions, number of voxels and peak MNI coordinates that have significantly different variances in brain response as well as the VS functional connectivity at T1 between the retained participants and retained plus attrited participants.

Regions	Voxel Number	MNI		
Brain Response				
Temporal_Inf_L	15	-33	-4	-32
Temporal_Inf_R	10	39	-4	-32
Frontal_Sup_L	6	-15	35	34
Left VS functional connectivity				
Fusiform_L	12	-36	-7	-20
Right VS functional connectivity				
Frontal_Mid_Orb_L	6	-39	53	-5
Fusiform_L	26	-36	-10	-32
Fusiform_R	7	39	-7	-32

As shown in **Table S5**, there was no significant difference in variances of the graph theory metrics between the retained participants and the retained plus attrited participants, suggesting excluding the attrited participants did not significantly change the variances of these measures.

Table S5. Difference in variances of the graph theory metrics between the retained participants and the retained plus attrited participants. The F values indicate the variance ratio between the two groups. The P values indicate if the ratio is significantly different from 1.

	F	***P***
Characteristic shortest path length	0.969666	0.646819
Strength	0.942205	0.374606
Clustering Coefficient	0.977784	0.738847

4. The use of the AUDIT cut scores seems potentially problematic. I commend the authors for considering normative development in creation of these cut scores and fully acknowledge that demarcation of substantive thresholds for adolescent substance use is extremely difficult. However, the cut scores used here have two clear weaknesses 1) there isn't clear prior literature support/discussion or outside data that informs these cut scores 2) no sensitivity analyses are performed to explore potential bias of these cut scores. Given the unique and extremely large sample of Imagen and what appears to be a fairly reasonable distribution of AUDIT scores at Time point 2, its unclear why a continuous metric (even with transformation [e.g., percentile if there are concerns about non-normality]) of change in AUDIT score was not explore.

Response:

(1) AUDIT cutoff

We have provided references supporting the cutoffs used for defining groups (see lines 458-472) and added the AUDIT distribution to the plot in Figure 6.

Materials and Methods

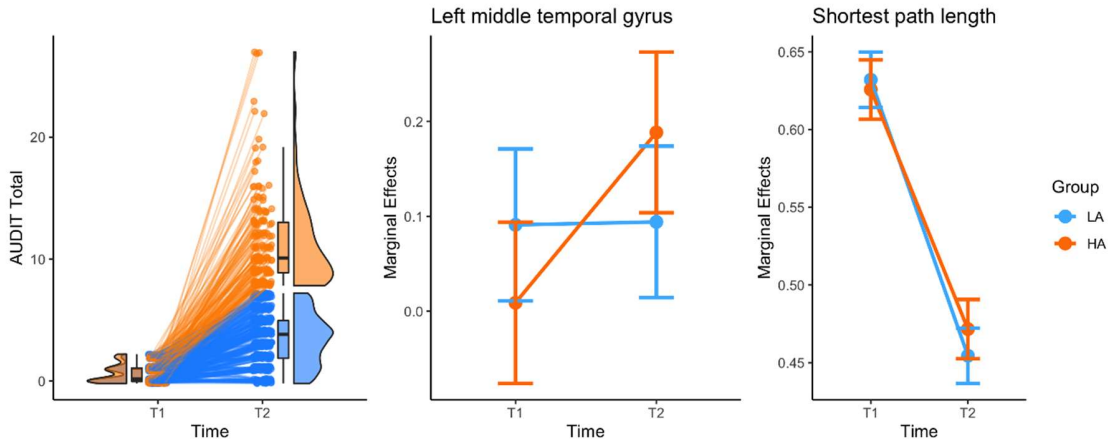
Alcohol use effects

A previous study showed that a cutoff of AUDIT ≥3 offered a good balance of sensitivity and specificity in detecting any alcohol use problems (sensitivity: 0.72; specificity: 0.89) as well as alcohol abuse or dependence (sensitivity: 0.88; specificity: 0.77) among adolescents (14-18 years) (Knight et al., 2003). Participants who reported an AUDIT total score ≥3 at T1 were excluded because they may have already experienced brain changes correlated with alcohol use. At T2 when alcohol use is more normative (the median AUDIT score was 6), participants who had an AUDIT total score ≤7 were designated as low-risk alcohol users and those who had an AUDIT total score ≥8 were considered as problematic drinkers. The cutoff used for T2 (≥8) is recommended for AUDIT in detecting problematic alcohol use in adults (Conigrave et al., 1995; WHO, 2001), which has shown relatively high sensitivity (0.82) and specificity (0.78-0.79) among young adults (18-25 years) (O'Connor et al., 2018). We defined those who had a low AUDIT score (≤2) at T1 but a high AUDIT score (≥8) at T2 as the high alcohol use group (HA). Those who had a low AUDIT score at both T1 (≤2) and T2 (≤7) were designated the low alcohol use group (LA).

Results

Alcohol use effects

Figure 6. Left: scatter plot of the AUDIT total scores at T1 and T2. Participants who had a low AUDIT score (⩽2) at T1 but a high AUDIT score (⩾8) at T2 were designated as the high alcohol use group (HA; orange). Those who had a low AUDIT score at both T1 (⩽2) and T2 (⩽7) were designated as the low alcohol use group (LA; blue). Each dot represents one participant with their T1 and T2 data connected by lines. Middle: The interaction results between time (T1 vs. T2) and alcohol group (LA vs. HA) for the functional connectivity between left middle temporal gyrus and right VS. To visualize this effect, functional connectivity within a spherical ROI with a 5-mm radius centered at the MNI coordinate of the left middle temporal gyrus was extracted and compared using the LME model. Right: The interaction results for the characteristic shortest path length.



(2) Sensitivity analyses

We have performed four sensitivity analyses to test if the observed alcohol effects with a T1 cutoff of ⩾3 and a T2 cutoff of ⩾8 replicate when:

(1). Using a different AUDIT cutoff (≥2) at T1.

(2). Using a variety of AUDIT cutoffs (≥6-13) at T2.

(3). Using the top and bottom 25th percentiles of T2 AUDIT to define HA and LA groups.

(4). Using change in AUDIT as a continuous measure instead of defining HA and LA groups.

The results of the sensitivity analyses indicated that splitting HA and LA based on the AUDIT cutoff (≥8) at T2 yielded reliable results, especially for the interactive effects of increased alcohol on the right VS connectivity with left middle temporal gyrus. The interaction effects remained significant in sensitivity analyses (1), (2) with cutoffs (⩾6-10), (3) and (4). The interaction effects for the shortest path length were replicated when using a different cutoff at T1 (≥2; i.e., participants with T1 AUDT ⩽1 were included) as well as using different cutoffs at T2 (≥7, ≥9). The interaction effects were non-significant in sensitivity analyses (3) and (4) but showed a same positive trend. It should be noted that the reduced sample size in HA group with higher T2 cutoffs in the sensitivity analysis (2) and in both HA and LA groups in the sensitivity analysis (3) can lower the statistical power to detect subtle effects. Examining brain effects at each level of the AUDIT would suffer from low statistical power. Aggregating group data according to a recommended AUDIT cutoff provides better insight with regard to subtle effects.

The sensitivity analyses have been included in the supplementary analysis section and the relevant information has been added to the manuscript (see lines 472-480; lines 587-598).

Materials and Methods

Alcohol use effects

… Those who had a low AUDIT score at both T1 (≤2) and T2 (≤7) were designated the low alcohol use group (LA). We performed four sensitivity analyses to test if the observed alcohol effects with a T1 cutoff of ⩾3 and a T2 cutoff of ⩾8 replicate when: (1) using a different AUDIT cutoff (≥2) at T1 (≥3 was used in the primary analysis); (2) using a variety of AUDIT cutoffs (≥6-13) at T2; (3) using top and bottom 25th percentiles for AUDIT at T2 to define HA and LA groups; (4) using change in AUDIT score as a continuous measure instead of defining HA and LA groups. Details can be found in the Supplementary Information…

Results

Alcohol use effects

The results of the sensitivity analyses indicated that splitting HA and LA based on the AUDIT cutoff (≥8) at T2 yielded reliable results, especially for the interactive effects of increased alcohol on the right VS connectivity with left middle temporal gyrus. The interaction effects remained significant in sensitivity analyses (1), (2) with cutoffs (≥6-10), (3) and (4). The interaction effects for the shortest path length were replicated when using a different cutoff at T1 (≥2) as well as using different cutoffs at T2 (≥7, ≥9). The interaction effects were non-significant in sensitivity analyses (3) and (4) but showed a same positive trend. It should be noted that the reduced sample size in HA group with higher T2 cutoffs in the sensitivity analysis (2) and in both HA and LA groups in the sensitivity analysis (3)

can lower the statistical power to detect subtle effects. Examining brain effects at each level of the AUDIT would suffer from low statistical power. Aggregating group data according to a recommended AUDIT cutoff may provide better insights with regard to subtle effects.

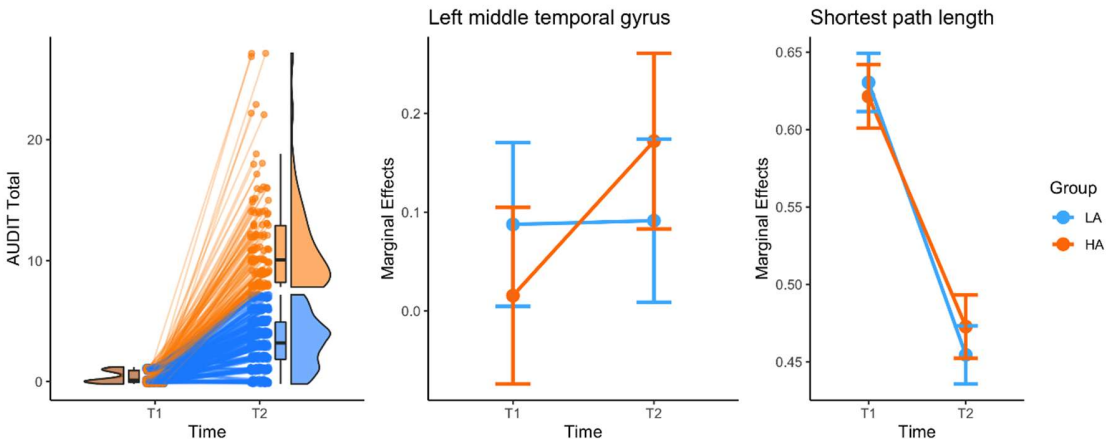
Supplementary Information

Sensitivity Analysis

We performed four sensitivity analyses to test if the observed alcohol effects with a T1 cutoff of ⩾3 and a T2 cutoff of ⩾8 replicate when: (1) using a different AUDIT cutoff (≥2) at T1; (2) using a variety of AUDIT cutoffs (≥6-13) at T2; (3) using top and bottom 25th percentiles for AUDIT at T2 to define HA and LA groups; (4) using change in AUDIT score as a continuous measure instead of defining HA and LA groups.

(1). We note that an alternative cutoff of AUDIT ≥2 offers a relatively good balance of sensitivity and specificity in detecting any alcohol use problems (sensitivity: 0.88; specificity: 0.81) as well as alcohol abuse or dependence (sensitivity: 0.93; specificity: 0.66) among adolescents (14-18 years) (Knight et al., 2003). We examined if a different cutoff at T1 would change the existing findings. The HA group was defined as participants who had a low AUDIT score (≤1) at T1 but a high AUDIT score (≥8) at T2. Those who had a low AUDIT score at both T1 (≤1) and T2 (≤7) were designated the low alcohol use group (LA). The results indicated that the findings were replicated with a stringent AUDIT cutoff (≥2) at T1. As shown in **Figure S7**, the positive interaction for functional connectivity between right VS and left middle temporal gyrus remained significant (t=3.75, 95% interval of null distribution: [-1.93,2.03]). The interaction for network strength network strength (t=-0.96, [-1.94, 1.99]) or network clustering coefficients (t=1.85, [-1.92, 1.97]) remained non-significant. The positive interaction remained significant on the characteristic shortest path length (t=2.67, [-1.93,1.94]; **Figure S7**).

Figure S7. Left: scatter plot of the AUDIT total scores at T1 and T2. Participants who had a low AUDIT score (⩽1) at T1 but a high AUDIT score (⩾8) at T2 were designated as the high alcohol use group (HA; orange). Those who had a low AUDIT score at both T1 (⩽1) and T2 (⩽7) were designated as the low alcohol use group (LA; blue). Each dot represents one participant with their T1 and T2 data connected by lines. Middle: The interaction results between time (T1 vs. T2) and alcohol group (LA vs. HA) for the functional connectivity between left middle temporal gyrus and right VS. To visualize this effect, functional connectivity within a spherical ROI with a 5-mm radius centered at the MNI coordinate of the left middle temporal gyrus was extracted and compared using the LME model. Right: The interaction results for the characteristic shortest path length.



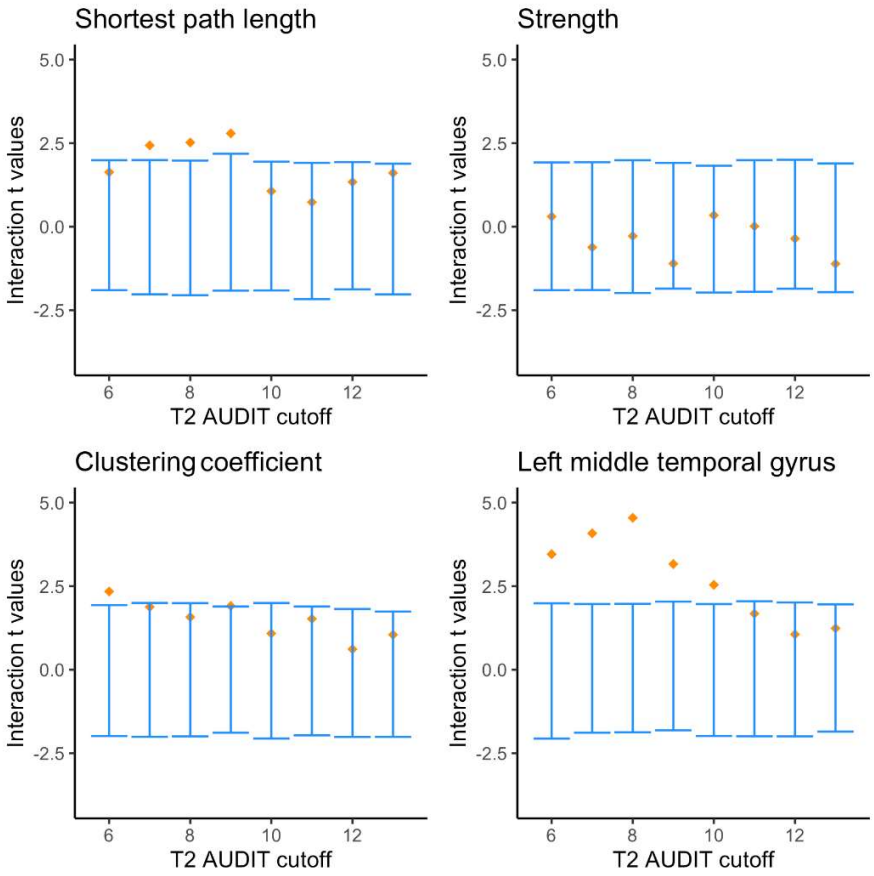
(2). We tested different AUDIT cutoffs (6-13) at T2. As shown in **Figure S8** and **Table S3**, the significant interaction effects reported in the primary analysis with a T2 AUDIT cutoff of ⩾8 were replicated with T2 AUDIT cutoffs of ≥7, ≥9. Notably, a higher cutoff resulted in fewer participants in the HA group, which may lower the statistical power to detect the interaction effects.

Table S3. Results of interactive effects between time (T1 vs. T2) and alcohol group (LA vs. HA) on network strength, shortest path length, clustering coefficient and the functional connectivity between left middle temporal gyrus and right VS. The alcohol group is defined using a variety of AUDIT cutoffs at T2. The last column indicates if the empirical t value exceeds the 95% interval of the null distribution.

T2 AUDIT cutoff	Interaction t values	Lower bound of 95% null distribution	Upper bound of 95% null distribution	Significant
Strength				
6	0.30	-1.90	1.92	No
7	-0.62	-1.90	1.93	No
8	-0.28	-1.98	1.99	No
9	-1.10	-1.85	1.91	No
10	0.34	-1.97	1.83	No
11	0.01	-1.95	1.99	No
12	-0.36	-1.86	2.00	No
13	-1.11	-1.96	1.89	No
Shortest path length				
6	1.63	-1.90	1.99	No
7	2.43	-2.02	1.99	Yes
8	2.52	-2.05	1.98	Yes
9	2.79	-1.91	2.18	Yes
10	1.06	-1.91	1.94	No
11	0.73	-2.17	1.91	No

12	1.34	-1.87	1.93	No
13	1.61	-2.02	1.88	No
Clustering coefficient				
6	2.34	-1.98	1.93	Yes
7	1.88	-2.01	2.00	No
8	1.58	-2.00	1.99	No
9	1.91	-1.88	1.89	Yes
10	1.08	-2.06	1.99	No
11	1.52	-1.97	1.89	No
12	0.61	-2.01	1.82	No
13	1.05	-2.01	1.74	No
Left middle temporal gyrus				
6	3.46	-2.06	1.99	Yes
7	4.08	-1.88	1.97	Yes
8	4.54	-1.87	1.97	Yes
9	3.16	-1.81	2.04	Yes
10	2.54	-1.98	1.97	Yes
11	1.68	-1.99	2.05	No
12	1.06	-1.99	2.01	No
13	1.24	-1.85	1.95	No

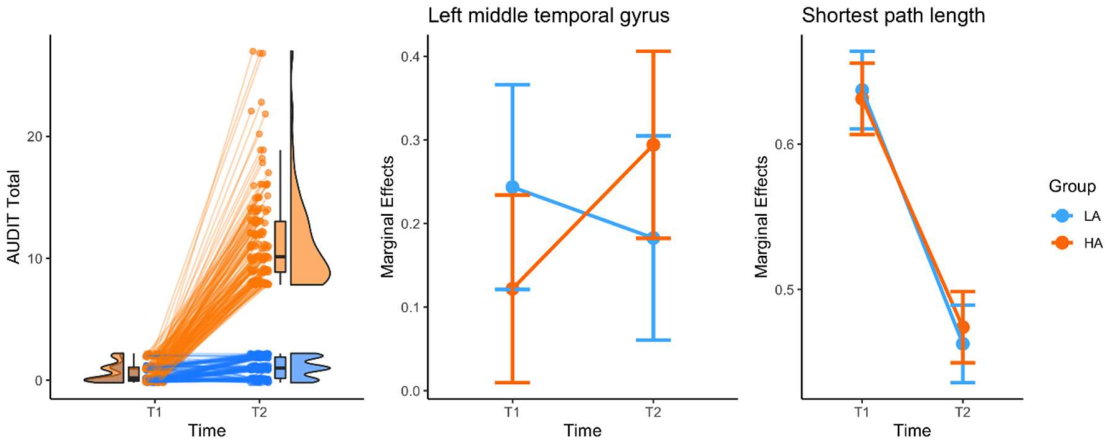
Figure S8. Results of interactive effects between time (T1 vs. T2) and alcohol group (LA vs. HA) on network strength, shortest path length, clustering coefficient and the functional connectivity between left middle temporal gyrus and right VS. The x-axis represents the AUDIT cutoff used to define the alcohol groups. The error bars represent the 95% interval of the null distribution, and the diamonds represent the empirical t values of the interaction term.



(3). As there is a trade-off between sensitivity and specificity when choosing a cutoff score for the AUDIT (e.g., a higher cut-off score may provide greater specificity but at the expense of sensitivity), we examined if a different strategy to defining groups at T2 would change the results. The HA group was defined as participants who had a low AUDIT score (≤2) at T1 but a high AUDIT score (top 25th percentile) at T2. Those who had a low AUDIT score at both T1 (≤2) and T2 (bottom 25th percentile) were designated the low alcohol use group (LA). As shown in **Figure S9**, the positive interaction for functional connectivity between right VS and left middle temporal gyrus remained significant (t=4.77, 95% interval of null distribution: [-2.04,1.92]). The interaction for network strength network strength (t=-0.14, [-1.88, 1.88]) or network clustering coefficients (t=1.71, [-1.98, 1.95]) remained non-significant. The interaction on the characteristic shortest path length showed a positive trend but did not reach significance (t=1.54, [-1.96,1.85]; **Figure S9**).

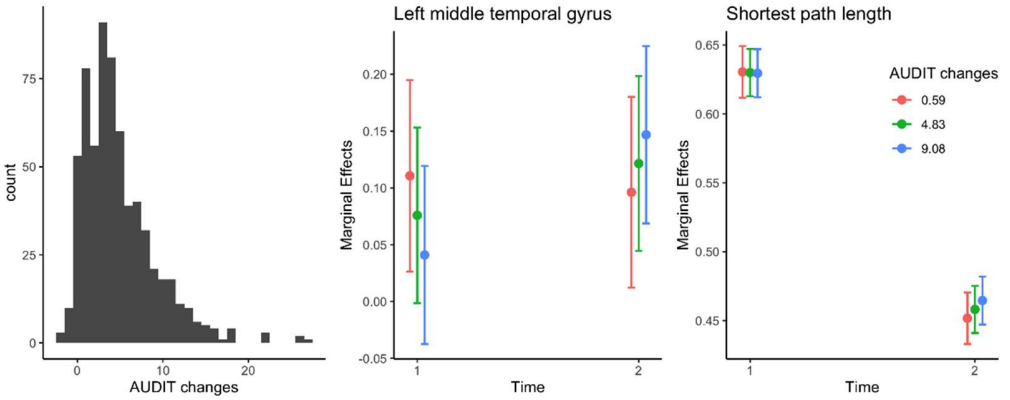
Figure S9. Left: scatter plot of the AUDIT total scores at T1 and T2. Participants who had a low AUDIT score (⩽2) at T1 but with a high AUDIT score (top 25th percentile) at T2 were designated the high alcohol use group (HA; orange). Those who had a low AUDIT score at both T1 (⩽2) and T2 (bottom 25th percentile) were designated as the low alcohol use group (LA; blue). Each dot represents one participant with their T1 and T2 data connected by lines. Middle: The interaction results between time (T1 vs. T2) and alcohol group (LA vs. HA) for the functional connectivity between left middle temporal gyrus and right VS. To visualize this effect, functional connectivity within a spherical ROI with a 5-mm radius centered at the MNI coordinate of the left middle temporal gyrus was extracted

and compared using the LME model. Right: The interaction results for the characteristic shortest path length.



(4). To explore if changes in AUDIT score across all participants interacted with the changes in reward-related functioning from T1 to T2, a separate LME model were performed with AUDIT changes from T1 to T2 for each participant included in the interaction term (time of the scan × AUDIT changes). As shown in **Figure S10**, the positive interaction for functional connectivity between right VS and left middle temporal gyrus remained significant (t=3.58, 95% interval of null distribution: [-1.99, 2.01]). The interaction for network strength network strength (t=-0.48, [-1.89, 1.96]) or network clustering coefficients (t=1.70, [-2.10, 1.96]) remained non-significant. No significant interaction was observed for network strength or network clustering coefficients. The interaction on the characteristic shortest path length showed a positive trend but did not reach significance (t=1.72, [-1.95,1.83]; **Figure S10**).

Figure S10. Left: histogram of the AUDIT changes from T1 to T2. Middle: results of interactive effects between time (T1 vs. T2) and alcohol changes on the functional connectivity between left middle temporal gyrus and right VS. Right: results of interactive effects on the characteristic shortest path length. The mean-value (4.83) and within one standard deviation (0.59, 9.08) of the AUDIT changes are used as grouping levels for the visualization.



5. Primary figures would be improved if some gradient of the effect were displayed. It appears these figures only display binarized masks of significant effects. More information would be useful to the reader, if for example the authors could display effect sizes within those voxels with significant effects. This is particularly useful, given the current state of neuroimaging research where there exists a few very large, consortium studies (IMAGEN included) with relatively high statistical power, but also many smaller, investigator-led samples with more modest sample sizes and statistical power. The display of effect sizes, rather than binarized statistical maps can help bridge the gap between these two types of studies, given the vast difference in statistical power.
Response:

We have calculated the effect sizes for the comparisons between T2 and T1 on brain response, VS connectivity. Relevant figures and texts have been updated in the manuscript (see lines 333-336; Figures 3 and 4).

Materials and Methods

Brain response and VS connectivity

The effect size of sampling time was calculated as

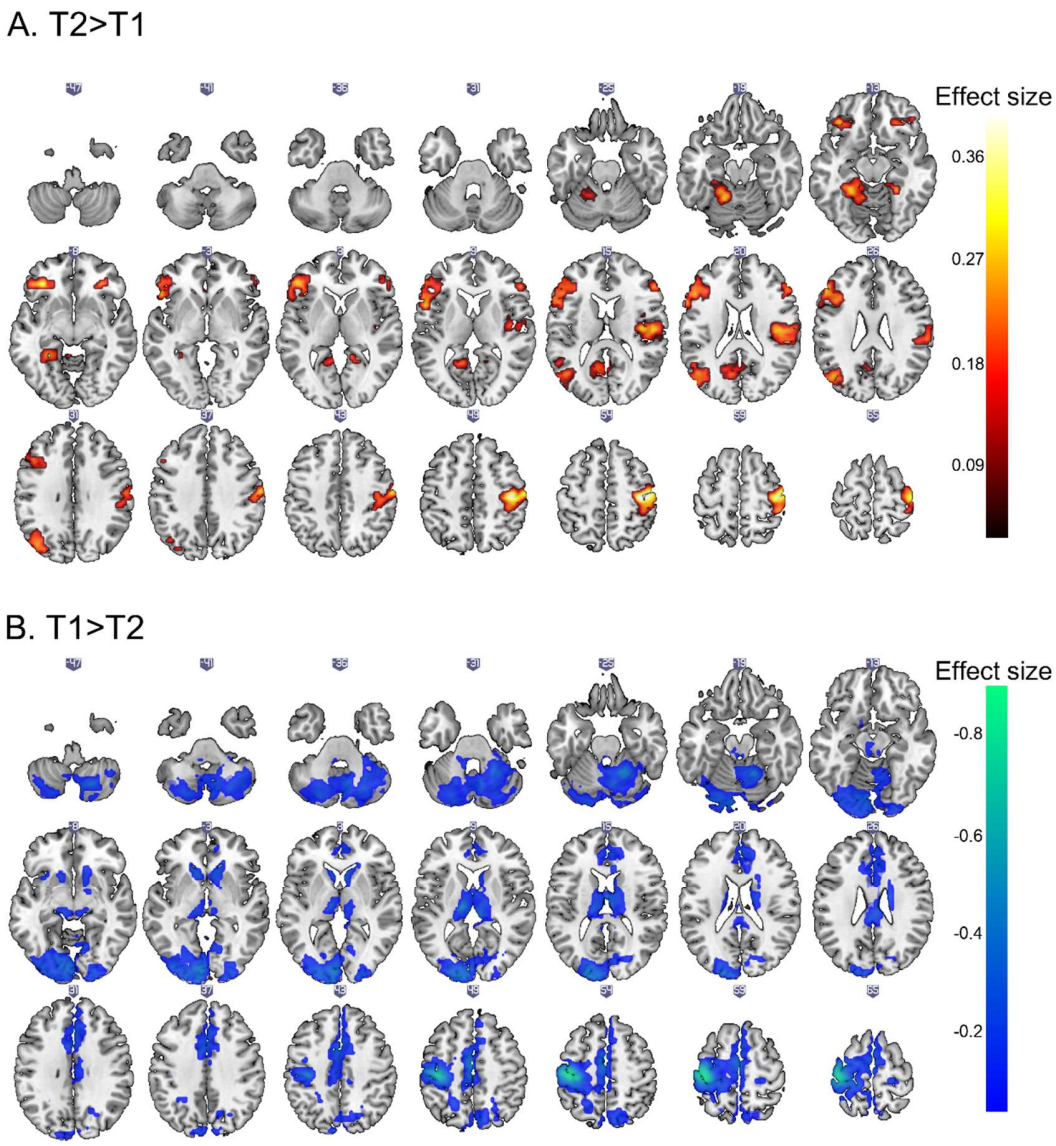
$$\text{Cohen's d} = \frac{t \times (\text{n1} + \text{n2})}{\sqrt{(n1 \times n2)} \times \sqrt{df}}$$

where n1 and n2 represent the numbers of participants at T1 and T2, respectively, and df is degrees of freedom.

Results

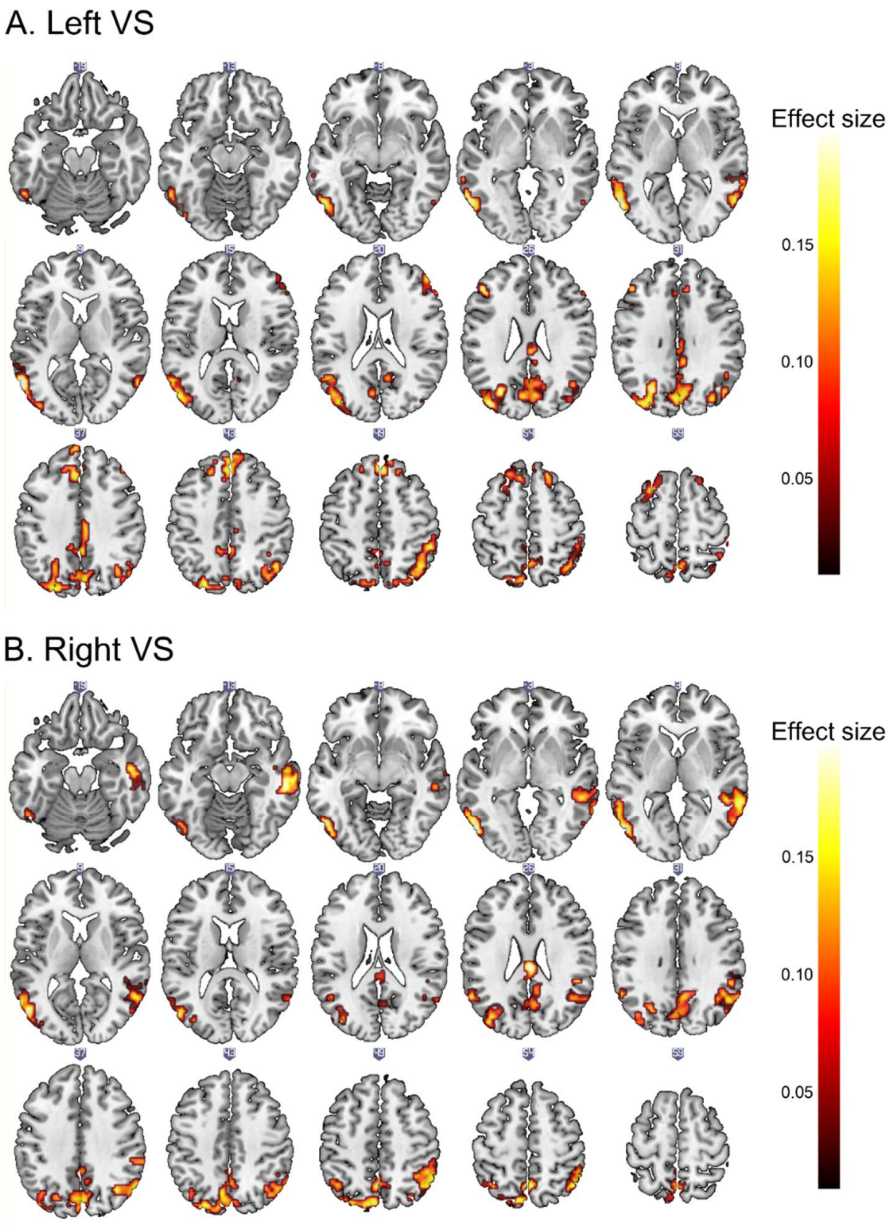
Brain response

Figure 3. Regions showing significantly increased (A) and decreased (B) brain response during reward anticipation at T2 as compared to T1. The colorbar denotes the effect size.



VS connectivity

Figure 4. Regions showing significantly increased functional connectivity with left (A) and right (B) VS during reward anticipation at T2 as compared to T1. The colorbar denotes the effect size.



Minor Comments:

Abstract:

* I would recommend clarifying the abbreviation timepoint for T1 the first time this is introduced, given the potential for T1 and T2 to be confused with MRI T1 versus T2 sequences.

* While potentially useful for those familiar with the AUDIT, the specific detail of AUDIT cut scores in the abstract will likely be difficult to parse for those outside of the substance use literature. It would be beneficial to additional include narrative descriptors, whatever those may be (but see additional comment regarding these cut scores) when listing these cut points in the abstract.

* "In all participants" may be confused with per-subject, idiographic/N=1 analysis. I would suggest this be revised to "Across the whole sample"

* In the abstract a mention of the contrast used in imaging analysis (i.e., large reward vs no-reward cue) would be useful, if space permits.

Response: We have addressed these issues in the abstract (see lines 135-162; lines 225-227).

> **Abstract**
>
> ... Participants from the IMAGEN study performed a Monetary Incentive Delay task during fMRI at timepoint 1 (T1; n=1,304, mean age=14.44 years old) and timepoint 2 (T2; n=1,241, mean age=19.09 years). The Alcohol Use Disorders Identification Test (AUDIT) was administrated at both T1 and T2 to assess a participant's alcohol use during the past year. Voxel-wise linear mixed effect models were used to compare whole brain response as well as functional connectivity of the ventral striatum (VS) during reward anticipation (large reward vs no-reward cue) between T1 and T2. In addition, task-modulated networks were constructed using generalized psychophysiological interaction analysis and summarized with graph theory metrics. To explore alcohol use in relation to development, participants with no/low alcohol use at T1 but increased alcohol use to hazardous use level at T2 (i.e., participants with AUDIT$\leqslant$2 at T1 and $\geqslant$8 at T2) were compared against those with consistently low scores (i.e., participants with AUDIT$\leqslant$2 at T1 and $\leqslant$7 at T2). Across the whole sample, lower brain response during reward anticipation was observed at T2 compared with T1 in bilateral caudate nucleus, VS, thalamus, midbrain, dorsal anterior cingulate as well as left precentral and postcentral gyrus…

> **Introduction**
>
> Here, participants from a large-scale longitudinal fMRI study (IMAGEN) performed the MID task during fMRI at timepoint 1 (T1; n=1,304, mean age=14.44 years old) and timepoint 2 (T2; n=1,241, mean age=19.09 years).

Introduction

* A careful review for typographical errors and general language would improve the manuscript

Response: We have carefully reviewed the introduction and revised the language issues.

* The term "efficiency" has received considerable criticism in the neuroimaging literature. It is worth considering whether more useful and detailed terminology can be used.

Response: The efficiency used in the manuscript refers to the topological efficiency (i.e., cost of transmitting information within the network; (Poldrack, 2015)). Network efficiency is a graph

theory metric that is defined as the inverse of shortest path length between each pair of nodes in the network. In the study, we used the characteristic shortest path length, that is the average number of shortest path lengths for all possible pairs of nodes in the network, to index how efficiently the network exchanges information. The potential vague use of “efficiency” has been changed to "network efficiency" or “topological efficiency”.

Highlights

Topological efficiency of task-modulated network increased at T2

Abstract

…Graph theory metrics of the task-modulated network showed higher inter-regional connectivity and topological efficiency at T2…

Discussion

Increased functional connectivity with VS and a more topologically efficient task-modulated functional network during reward anticipation were found.

...could be reflected in the lower functional connectivity and topological efficiency of the brain’s reward network that we observed at T1…

Conclusion

Mirroring these regional changes, the functional connectivity with the VS increased and the brain network involved in reward anticipation also became more topologically efficient.

Methods

* +- should be defined at least in the first instance in the methods section.

Response: We have addressed this comment (see lines 255-258).

After the quality check on neuroimaging and behavioral data, there were 1304 participants (female: 681; mean and standard deviation age: 14.44±0.41 years; modal Pubertal Development Scale pubertal development status: female = 4-advaned pubertal, male=3-midpubertal) at T1 and 1241 participants (female: 639, mean and standard deviation age: 19.09±0.76 years) at T2 included in the study.

* Pubertal scores, while potentially helpful for the demographic tables, 1) likely don't require detail description in methods since they only appear to be used as sample descriptions and 2) the values themselves are appear confusing. The distribution of the PDS appears not to change from T1 to T2 even though 5 years have gone by? I would expect some beginning puberty scores at 14, but that the same amount occurs when the cohort is ~19 seems like measurement error.

Response: The baseline puberty stages were included in the LME models. Participants were assumed to be post-pubertal at T2. As the participants were not matched between T1 and T2 in the first set of analysis, we also reported the baseline pubertal stages for those included in the T2. We have clarified this point in Table 1.

Table 1. Characteristics of participants at T1 and T2.

	T1 (N = 1304)	T2 (N = 1241)	*p*
Sex, n (%)			0.74
Male	623 (48)	602 (49)	
Female	681 (52)	639 (51)	
Age, years			
Mean ± SD	14.44 ± 0.41	19.09 ± 0.76	
Handedness, n (%)			0.63
Left	141 (11)	126 (10)	
Right	1,163 (89)	1,115 (90)	
Mean FD			
Mean ± SD	0.25 ± 0.17	0.17 ± 0.10	<0.05
T1 PDS*, n (%)			0.95
Pre-pubertal(1)	9 (1)	8 (1)	
Beginning pubertal(2)	73 (6)	76 (6)	
Mid-pubertal(3)	399 (31)	365 (29)	
Advanced pubertal(4)	744 (57)	719 (58)	
Post-pubertal(5)	79 (6)	73 (6)	
Left/Right response ratio			
Mean ± SD	1.07 ± 0.31	1.09 ± 0.33	0.21
RT (ms), Mean ± SD			
Large reward	246.35 ± 31.75	228.70 ± 32.02	
Small reward	254.84 ± 36.44	234.42 ± 34.43	
No reward	278.77 ± 45.75	249.39 ± 40.48	
Large-No reward	-32.42 ± 33.87	-20.69 ± 29.86	

FD: Framewise displacement; RT: Response time; PDS: Pubertal development stage; SD: standard deviation. Two sample t-test was used to compare mean FD and chi-square tests were used to compare sex, handedness and T1 PDS respectively between T1 and T2. * Participants are assumed to be post-pubertal at T2. As the participants are not matched between T1 and T2, the T1 PDS is also reported for participants at T2.

Reviewer #2

This is a well-written paper that holds potential to illuminate developmental differences in reward-anticipation-related connectivity with other brain regions, using the very large IMAGEN sample.

My chief concern about this paper in its present form is that its framing is solely within the opponent-process story, where it is assumed that VS activation is all hedonics. There is an entire literature showing that VS can be recruited solely by vigilance or other cognitive demand, even in the absence of reward. c.f.

Sarter M, Gehring WJ, Kozak R. More attention must be paid: the neurobiology of attentional effort. Brain research reviews. Aug 2006;51(2):145-160.

Breckel TP, Giessing C, Thiel CM. Impact of brain networks involved in vigilance on processing irrelevant visual motion. NeuroImage. Apr 15 2011;55(4):1754-1762.

Boehler CN, Hopf JM, Krebs RM, Stoppel CM, Schoenfeld MA, Heinze HJ, Noesselt T. Task-load-dependent activation of dopaminergic midbrain areas in the absence of reward. The Journal of neuroscience : the official journal of the Society for Neuroscience. Mar 30 2011;31(13):4955-4961.

Therefore, a possibility exists that VS activation is a combination of activation by anticipatory AFFECT (delight at reward prospects) and mobilization of attentional/cognitive resources. For example, adolescent VS activation may stem more from the former, and adult activation from the latter. Notably, reaction times shortened across the board at T2.

Here you have a very appropriate means of directly probing the degree to which reward anticipation shows functional linkages to these other frontocortical regions that govern task maintenance and attention! This would be a much more solid paper if the intro was framed in that literature and if the findings were explained (e.g. in the discussion) in terms of this other literature on the non-hedonic components of reward-directed anticipatory VS recruitment.

Response:

We appreciate the insightful comments. We have addressed this comment in the introduction (see lines 202-207) and discussion (see lines 639-654).

Introduction

.. For instance, we have previously reported connectivity between bilateral VS and regions that play a role in attention to and integration of salient information (e.g., middle, inferior frontal gyrus, angular gyrus, inferior parietal gyrus, insula and putamen) during reward anticipation in adolescents (Cao et al., 2019). Thus, an examination of VS connectivity can probe the involvement of the non-hedonic components that are linked to the VS activity such as mobilization of attentional or cognitive resources.

Discussion

...It is possible that the right hemisphere specific response in the motor area is a characteristic of anticipation of large reward in contrast to no reward in adults.

A body of literature has shown that reward-related regions such as the VS can be recruited not only for reward-specific processes but also in the service of vigilance or other cognitive demands, even in the absence of reward (Sarter et al., 2006; Boehler et al., 2011; Breckel et al., 2011; Vassena et al., 2014). For instance, Boehler and colleagues reported that participants had greater VS response in a high-demand cognitive task when compared to a low-demand task, suggesting the VS is also involved in the recruitment of processing resources even in the absence of monetary reward (Boehler et al., 2011). It is possible that the VS response observed during reward anticipation may reflect a combination of response to anticipatory affect (i.e., delight at reward prospects) and mobilization of

attentional or cognitive resources. Thus, decreased VS response from T1 to T2 would be consistent with the adolescent VS response stemming more from anticipatory affect while the adult response may stem more from non-hedonic processes such as mobilization of attentional or cognitive resources. In support of this idea, we found age-related increases in VS functional connectivity with frontoparietal regions such as the inferior frontal gyrus, parietal gyrus, middle and inferior occipital gyrus which play important roles in value-driven attention capture and in controlling attentional resources (Corbetta and Shulman, 2002; Anderson, 2016).

Also, while the authors laudably acknowledge how the reward task (MID) used candies at baseline and some abstract exchange for cash with unrevealed payouts at T2. However, this could be a critical factor in explaining differences from other findings. There is a large literature on how VS activation scales with explicit signaling of actual monetary rewards (PUBMED 23507394). To assume that you have found the "correct" directionality of developmental differences is not justified. Perhaps this directionality is specific to abstract or mysterious rewards and would not extend to concrete (and significant on a single- trial-level) rewards.

Response:

The points-to-reward ratio was explicit at T1 (5 points = 1 candy) but was uncertain at T2. The reviewer raises the concern that the added uncertainty element at T2 may be a confounding factor. We suspect this is unlikely to be the case for a number of reasons.

The cross-sectional analysis using a probabilistic mapping approach indicated that the key regions involved in reward anticipation largely overlapped between T1 and T2. If the added uncertainty element is consequential, we might predict the recruitment of additional regions that encode uncertainty (e.g., OFC/mPFC). A large number of studies have shown that uncertainty signals are encoded by OFC (Hsu et al., 2005; Knutson et al., 2005; Schultz et al., 2008; Abler et al., 2009). For example, ambiguous options with incomplete probabilistic information elicited stronger brain signals than options with complete probabilistic information in OFC (Schultz et al., 2008). In support of this, patients with OFC lesions were insensitive to the level of uncertainty in behavioral choices (Hsu et al., 2005). Further, brain response in OFC were related to reward uncertainty when participants anticipated potential monetary rewards with varying levels of uncertainty (Abler et al., 2009). Notably, a previous study used a variant of the standard MID task where probabilities (i.e., level of uncertainty) and magnitude of the reward were signaled by different types of cues. This study reported that that the VS response was associated with anticipated gain magnitude while the anticipated gain probability was associated with brain response in the mesial prefrontal cortex (Knutson et al., 2005). However, in the present study we found neither a robust response in OFC/mPFC during reward anticipation at T2 nor were there significant differences in brain response in OFC/mPFC between T1 and T2. This may suggest that the added uncertainty element was not explicitly processed or attended to during task performance, due perhaps to the relatively short interval for cue display and reward anticipation preventing participants from a deliberation on the probability of winning the reward. Furthermore, it may be important to note that the uncertainty in the present task may be different in important ways to the uncertainty that is typically researched. Typically, experimental manipulations of uncertainty add a probabilistic element to whether or not a successful response will be rewarded. In contrast, in the T2 task, participants knew that all successful responses would be rewarded and there was no uncertainty

that a successful response would yield points won – what was unknown was the value of the points.

The instructions given to participants at both T1 and T2 instructed them to win as many points as possible. The RT data showed the anticipated reward effects (i.e., participants responded faster to large rewards than no rewards) at both T1 and T2 indicating that they were motivated to perform well, independent of how the points were exchanged at the end of the experiment.

We have provided more information regarding the instructions given to participants in the method section (see lines 271-273) and included a discussion regarding the added uncertainty element at T2 (see lines 718-744).

Materials and Methods

Experimental Design

At T1, there were 66 trials in total and 22 trials per condition. Participants were given an M&M for every 5 points they won. At T2, there were 42 trials in total and 14 trials per condition. Participants were instructed that they could exchange their points for money at the end to increase their motivation. Every participant, independent of their performance, received £/€5 (cash or vouchers). At both time points, participants were asked to win as many points as possible. The instructions given to participants at T1 and T2 are provided in the Supplementary Information.

Discussion

A few limitations need to be borne in mind when interpreting the results. Due to ethical concerns, participants were given candies at T1. Different reward types between T1 and T2 could be a potential confounding factor. Although equating rewards across large developmental periods is always a significant challenge, we note that the behavioural measures (i.e., faster responding for large rewards compared to no rewards) indicated that participants were motivated by the rewards available at both T1 and T2 regardless of how the points were exchanged at the end of the experiment. Another limitation is that the points-to-reward ratio was explicit at T1 (5 points = 1 candy) but was uncertain at T2. However, we suspect that the added uncertainty element at T2 was unlikely to be consequential. The cross-sectional analysis using a probabilistic mapping approach indicated that the key regions involved in reward anticipation largely overlapped between T1 and T2. If the added uncertainty element is consequential, we might predict the recruitment of additional regions that encode uncertainty (e.g., OFC/mPFC). A large number of studies have shown that uncertainty signals are encoded by OFC (Hsu et al., 2005; Knutson et al., 2005; Schultz et al., 2008; Abler et al., 2009). For example, a previous study used a variant of the standard MID task where probabilities (i.e., level of uncertainty) and magnitude of the reward were signaled by different types of cues. This study reported that that the VS response was associated with anticipated gain magnitude while the anticipated gain probability was associated

with brain response in the mesial prefrontal cortex (Knutson et al., 2005). However, in the present study we found neither a robust response in OFC/mPFC during reward anticipation at T2 nor were there significant differences in brain response in OFC/mPFC between T1 and T2. This may suggest that the added uncertainty element was not explicitly processed or attended to during task performance, due perhaps to the relatively short interval for cue display and reward anticipation preventing participants from a deliberation on the probability of winning the reward. Furthermore, it may be important to note that the uncertainty in the present task may be different in important ways to the uncertainty that is typically researched. Typically, experimental manipulations of uncertainty add a probabilistic element to whether or not a successful response will be rewarded. In contrast, in the T2 task, participants knew that all successful responses would be rewarded and there was no uncertainty that a successful response would yield points won – what was unknown was the value of the points. Moreover, multiple factors such as novelty seeking, family history of drug abuse, genetic variance have been shown related to lower VS response during reward anticipation in previous studies (Büchel et al., 2017; Maričić et al., 2020; Tschorn et al., 2021), which need to be considered in future studies.

Supplementary Information

Experiment instruction at T1

This task is a reaction time task - it tests how quickly you can press the button to hit a target, which is a white square appearing only for a short time on the left or right of the screen. If you manage to press the button as soon as the white square appears, you will score points. If you respond too early (before the white square appears) or too late (after the white square has disappeared) you will not gain any points. You can tell where the white square will appear and how many points you will win by the symbol you see on the screen before the white square is shown. A triangular symbol means you will not win any points, a circle with a line means you will win 2 points and a circle with three lines means you will win 10 points. You should try to win as many points as you can! - but only if you press the button while the square is presented on the screen! For every 5 points you win, you will receive an M&M, -see how many you can win!

Experiment instruction at T2

This task is a reaction time task - it tests how quickly you can press the button to hit a target, which is a white square appearing only for a short time on the left or right of the screen. If you manage to press the button as soon as the white square appears, you will score points. If you respond too early (before the white square appears) or too late (after the white square has disappeared) you will not gain any points. You can tell where the white square will appear and how many points you will win by the symbol you see on the screen before the white square is shown. A

triangular symbol means you will not win any points, a circle with a line means you will win 2 points and a circle with three lines means you will win 10 points. You should try to win as many points as you can! - but only if you press the button while the square is presented on the screen! Your points will be exchanged for cash, let's see how much you can win!

Notes for instructors: The subjects will exchange their points for money (cash or vouchers) at the end to increase their motivation. Every participant, independently of their performance, will receive £/€ 5 (however, they must not be told this! They should believe that the final gain is related to their performance.)

Reference

Abler B, Herrnberger B, Grön G, Spitzer M (2009) From uncertainty to reward: BOLD characteristics differentiate signaling pathways. BMC neuroscience 10:1-12.

Anderson BA (2016) The attention habit: How reward learning shapes attentional selection. Annals of the new York Academy of Sciences 1369:24-39.

Bjork JM, Smith AR, Chen G, Hommer DW (2010) Adolescents, adults and rewards: comparing motivational neurocircuitry recruitment using fMRI. PloS one 5:e11440.

Boehler CN, Hopf J-M, Krebs RM, Stoppel CM, Schoenfeld MA, Heinze H-J, Noesselt T (2011) Task-load-dependent activation of dopaminergic midbrain areas in the absence of reward. Journal of Neuroscience 31:4955-4961.

Breckel TP, Giessing C, Thiel CM (2011) Impact of brain networks involved in vigilance on processing irrelevant visual motion. Neuroimage 55:1754-1762.

Büchel C, Peters J, Banaschewski T, Bokde AL, Bromberg U, Conrod PJ, Flor H, Papadopoulos D, Garavan H, Gowland P (2017) Blunted ventral striatal responses to anticipated rewards foreshadow problematic drug use in novelty-seeking adolescents. Nature communications 8:1-11.

Cao Z, Bennett M, Orr C, Icke I, Banaschewski T, Barker GJ, Bokde AL, Bromberg U, Büchel C, Quinlan EB (2019) Mapping adolescent reward anticipation, receipt, and prediction error during the monetary incentive delay task. Human brain mapping 40:262-283.

Conigrave KM, Hall WD, Saunders JB (1995) The AUDIT questionnaire: choosing a cut-off score. Addiction 90:1349-1356.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nature reviews neuroscience 3:201-215.

Di X, Biswal BB (2019) Toward task connectomics: examining whole-brain task modulated connectivity in different task domains. Cerebral Cortex 29:1572-1583.

Ernst M, Torrisi S, Balderston N, Grillon C, Hale EA (2015) fMRI functional connectivity applied to adolescent neurodevelopment. Annual review of clinical psychology 11:361-377.

Fair DA, Dosenbach NU, Church JA, Cohen AL, Brahmbhatt S, Miezin FM, Barch DM, Raichle ME, Petersen SE, Schlaggar BL (2007) Development of distinct control networks through segregation and integration. Proceedings of the National Academy of Sciences 104:13507-13512.

Finn ES (2021) Is it time to put rest to rest? Trends in Cognitive Sciences.

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680-1683.

Knight JR, Sherritt L, Harris SK, Gates EC, Chang G (2003) Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. Alcoholism: Clinical and experimental research 27:67-73.

Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. Journal of Neuroscience 25:4806-4812.

Lorenz RC, Gleich T, Beck A, Pöhland L, Raufelder D, Sommer W, Rapp MA, Kühn S, Gallinat J (2014) Reward anticipation in the adolescent and aging brain. Human brain mapping 35:5153-5165.

Lv H, Wang Z, Tong E, Williams LM, Zaharchuk G, Zeineh M, Goldstein-Piekarski AN, Ball TM, Liao C, Wintermark M (2018) Resting-state functional MRI: everything that nonexperts have always wanted to know. American Journal of Neuroradiology 39:1390-1399.

Maričić LM, Walter H, Rosenthal A, Ripke S, Quinlan EB, Banaschewski T, Barker GJ, Bokde AL, Bromberg U, Büchel C (2020) The IMAGEN study: a decade of imaging genetics in adolescents. Molecular Psychiatry 25:2648-2671.

Mennes M, Kelly C, Colcombe S, Castellanos FX, Milham MP (2013) The extrinsic and intrinsic functional architectures of the human brain are not equivalent. Cerebral cortex 23:223-229.

O'Connor EA, Perdue LA, Senger CA, Rushkin M, Patnode CD, Bean SI, Jonas DE (2018) Screening and behavioral counseling interventions to reduce unhealthy alcohol use in adolescents and adults: updated evidence report and systematic review for the US Preventive Services Task Force. Jama 320:1910-1928.

Poldrack RA (2015) Is "efficiency" a useful concept in cognitive neuroscience? Developmental cognitive neuroscience 11:12-17.

Sarter M, Gehring WJ, Kozak R (2006) More attention must be paid: the neurobiology of attentional effort. Brain research reviews 51:145-160.

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo CD, Tobler PN, Bossaerts P (2008) Explicit neural signals reflecting reward uncertainty. Philosophical Transactions of the Royal Society B: Biological Sciences 363:3801-3811.

Stevens MC (2016) The contributions of resting state and task-based functional connectivity studies to our understanding of adolescent brain network maturation. Neuroscience & Biobehavioral Reviews 70:13-32.

Tschorn M, Lorenz RC, O'Reilly PF, Reichenberg A, Banaschewski T, Bokde AL, Quinlan EB, Desrivières S, Flor H, Grigis A (2021) Differential predictors for alcohol use in adolescents as a function of familial risk. Translational psychiatry 11:1-11.

Vassena E, Silvetti M, Boehler CN, Achten E, Fias W, Verguts T (2014) Overlapping neural systems represent cognitive effort and reward anticipation. PLoS One 9:e91008.

WHO (2001) AUDIT: The alcohol use disorders identification test: Guidelines for use in primary health care. In: World Health Organization.