Social cognition as a matter of structural brain connections: a systematic review and diffusion weighted imaging meta-analysis

Rita Hansl, Lara Z. Maliske, Sofie Valk, Philipp Kanske

OSF project folder: https://osf.io/3z4bf/?view_only=ca95cb2546604b6ab7da562fbee68d39
Stage 1 manuscript: https://osf.io/t9wfg?view_only=ca95cb2546604b6ab7da562fbee68d39
ArticleID #907
version 1

1. Recommender: Marietta Papadatou-Pastou

Recommender comments	Author response
a) I have now received evaluations of your Stage 1 submission from two expert reviewers. Based on these comments, we cannot accept your manuscript in its present form but would like to invite you to revise your article, addressing the issues raised by the reviewers and myself below: Additionally, I have a few concerns and suggestions of my own:	Thank you very much for the smooth handling process and the seemingly instant and most constructive feedback. We have considered all the valuable and helpful input and tried to clarify and incorporate everything to the best of our abilities. We appreciate the chance to further improve our project and look forward to your response.
b) 1. Please ensure that paragraphs are shorter than one page (when double-spaced, 12-point font), in accordance with APA style.	Thank you for this remark. We implemented the respective changes throughout the manuscript.
c) 2. For clarity, I recommend rephrasing the following:	Thank you for these concise and constructive suggestions.
d) Page 8: "Given that neither co-varying activity implies a direct neural connection or interaction, researchers have highlighted the value of structural connectivity as a measure of functional brain organization (e.g., Forkel et al., 2022)."	The sentence in section **1.2 Structural connectivity** was rephrased to: However, co-varying activity neither implies a direct neural connection nor interaction. Therefore, researchers have highlighted the value of structural connectivity as a measure of functional brain organization (e.g., Forkel et al., 2022)
e) Page 12: "MA1) To examine the overall relationship between metric measures of social cognition and structural connectivity (RA1), correlations in all identified studies are meta-analyzed across socio-cognitive constructs, DTI metrics, populations/diagnoses, and methodologies."	The sentence in section **2.2. Design** was reformulated: **MA1)** To examine the overall relationship between metric measures of social cognition and structural connectivity (**RA1**) correlations in all identified studies are meta-analyzed. Thereby, studies investigating different *socio-cognitive constructs*, *DTI-metrics, populations/diagnoses* and *methodologies* are integrated and the study variability is accounted for using moderator analysis.
f) 3. On Page 9, you mention "the benefit of novel, more fine-grained analysis techniques." Could you please specify which techniques these are?	We apologize for not being more clear and are happy to expand on this point in more detail. More fine-grained measures as discussed in the cited papers include: fixel-

	based analysis, bundle analytics, or advanced "multidimensional" diffusion MRI acquisitions (e.g. Chandio et al., 2020; Dhollander et al., 2021; O'Donnell et al., 2019). This information has been added in-text in the last paragraph of section **1.2. Structural connectivity**. For more detailed information we kindly refer to the provided references.
g) 4. The authors discuss "transdiagnostic integration." By grouping all diagnoses together, I wonder if some information is lost. Could diagnosis also be used as a moderator (if sufficient data points are available)? Comparisons between different diagnoses would be valuable.	We thank you for raising this very important point! We agree that diagnosis is likely to be an important moderating factor which is why *population/diagnosis* has been integrated as a variable in our planned models. We understand that the phrasing might have been misleading since the aim was and is to use the nominal (not binary) variable *population/diagnosis* to differentiation between different diagnoses. We have clarified this in the list of moderators on page 11 as well as the model specifications.
h) 5. Regarding Exclusion Criterion 10: "not having undergone peer review except for primary data," this phrasing is unclear. It seems contradictory to require only peer-reviewed data while including analyses of primary data that have not been peer-reviewed (unless I am misunderstanding something). Additionally, how will the primary data be collected? Will the authors contact the authors of published studies, reach out to researchers in the field, or identify any databases? If primary data is involved, will the authors perform the analyses themselves, and what analytical decisions will they make?	Thank you for highlighting this unclarity and allowing us to elaborate on our intentions. Indeed, with this point we describe our aspirations to request more detailed data from authors of included peer-reviewed studies, rather than calling for unpublished data. The idea is to contact the authors of included papers and inquire about potential additional statistical maps that are not available within the encountered publications. This point only concerns group-level statistical maps for inclusion in the whole-brain neuroimaging meta-analysis. Therefore, no further analysis of unpublished primary data will be necessary. This information has been added at the end of section **2.3 Search strategy.**
i) 6. The list of excluded studies, along with reasons for exclusion, should also be provided as supplementary material, and this intention should be stated in the preregistration.	Thank you for this important remark. We intend to list all screened articles in the coding sheet which should make the exclusion most transparent. However, we are happy to provide an additional list of excluded articles if that allows for a clearer overview. This information has been additionally stated at the end of section **2.4 Inclusion and exclusion criteria.**
j) 7. The search strategy could be enhanced by checking the citations of included studies, as well as forward-searching studies that cite the included studies. Moreover, consider searching the reference lists of important reviews in the area (e.g., Wang et al., 2018).	Thank you for this valuable suggestion. We extended our search strategy by forward- and backward-searches, as suggested, and incorporated this adaptation in the methods section.

k)	8. In Figure 1, one reason for removing records before screening is "records marked as ineligible by automation tools." This is not described in the text, and I am unsure what it means. Will the authors not check the eligibility of those records as well? Additionally, in the same cell, "Records removed for other reasons"—what reasons other than duplicates might there be before the actual screening phase? Please clarify.	Thank you for the thorough review. We have adapted the template to better fit the planned and described procedure.
l)	9. Why is sex ratio used as a moderator instead of comparing data from the two sexes? Perhaps the authors should first determine if enough studies have broken down data by sex; if not, then using the sex ratio as a moderator would be appropriate.	Thank you for this remark, and we are happy to clarify our approach here. We expect studies to report basic sample characteristics which would include participants' sex. If not reported, this information should also be retrievable from the authors. We therefore think it will be feasible to include the sex-ratio within each sample as a moderator. However, we do not intend to dissect samples by sex, nor to perform direct sex comparisons since we do not expect studies to report effect sizes separately by sex.
m)	10. On Page 17, you state, "For MA2 and MA3, additional meta-analyses are calculated for socio-cognitive constructs and DTI metrics analyzed by a minimum of 5 studies to gain more thorough insights into interactions and moderation." Will these be separate meta-analyses for each level of the moderator, or will a typical moderating analysis be conducted? The latter should be preferred, as it allows for statistical comparisons between levels. If analyses are separate, they should not be termed "moderating analyses," and the variables should not be referred to as "moderators." Section 2.7.2.1 refers to these as sub-group analyses, but "separate analyses" are mentioned elsewhere. This distinction should be clarified.	Thank you for raising this important point! According to the current plan, we intend to perform one main analysis including all studies, where moderating effects will be analyzed using meta-regression and subgroup analysis. Additionally, we want to perform separate meta-analyses of socio-cognitive constructs and DTI metric, to see if results hold in more homogenous data and to be able to investigate interactions more thoroughly. For example, if there were more than 5 studies using FA as the DTI-metric of interest, this more homogenous sub-sample data would be used for a meta-analysis including meta-regression and subgroup analysis of the remaining moderators. We hope for this approach to provide more thorough insights into the interactions and potential construct or metric specific effects. We understand that the description of the planned subgroup analysis was unclear and have revised the use of the term moderator and modified section ***2.7.2 Meta-regression and sub-group analysis***. We hope our efforts were successful.
n)	11. Since a Bayesian framework is used, it would be advisable to perform some robustness checks. In addition to the chosen prior, a range of other priors should also be considered to see if the results differ. The aim is to determine how stable the inferences drawn from the	We understand and generally share your concerns regarding model robustness. The chosen Bayesian framework, named *Robust Bayesian Meta-analysis*, is designed to be more robust than other methods by averaging over a range of different models that address publication bias in different

model are under various scenarios or settings. This is crucial, as Bayesian analysis often involves subjective choices in its priors, which can significantly influence outcomes.	ways. The results of the underlying models will be reported as proposed by the developers (Maier et al., 2023). The visualization of the output for the model comparison can be found within the R package documentation: https://fbartos.github.io/RoBMA/ Since the aim is to perform confirmatory hypotheses testing, we would refrain from defining additional sets of priors a priori but adhere to the developers' suggested configurations (Bartoš et al., 2023). However, since we do agree to the importance of testing robustness, post-hoc assessments of alternative prior configurations will be performed using a novel tool by Höfler (2021) (Bayesian Regions of Evidence). This approach shall allow sensitivity analysis and the impact of prior selection. Michael Höfler, "Bayesian Regions of Evidence (for Normal Distributions)" (OSF, October 28, 2021), https://doi.org/10.31234/osf.io/mg23h.
o) M12. In Table 1, last cell, rows 1 & 3: "Strong evidence against H1 would indicate a lack of the hypothesized correlation between structural connectivity and socio-cognitive abilities." This should be rephrased using Bayesian terminology, for example, "strong evidence in favor of the null hypothesis." In Table 1, Hypothesis 2: "Prior evidence can be integrated into brain maps identifying the areas where diffusion metrics most strongly correlate with socio-cognitive functions." This statement does not seem to be phrased as a hypothesis.	Thank you for the remark. The suggestion was implemented in the study design table at the end of the document. H2 has be rephrased to: Associations between structural connectivity and socio-cognitive functions are localized in specific brain regions.
p) Thank you for considering my feedback. I look forward to your thoughts!	Thank you for the concrete and constructive input! We hope we could successfully implement your valuable feedback.
q) Best regards, Marietta Papadatou-Pastou	

2. Reviewer 1: Sebastian Ocklenburg, 27 Nov 2024 14:26

Reviewer comments	Author's response
a) Review of Stage 1 RR "Meta-analysis: Social cognition and structural connectivity"	

Predefined criteria:	
b) 1A. The scientific validity of the research question(s). The three research aims stated in section 1.4 all have high scientific validity and the introduction makes it clear why it makes sense to investigate these aims.	
c) 1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable. While section 1.4 is named "Research aims and hypotheses" it actually does not contain any hypotheses. I would like to encourage the authors to provide clear, directional and testable hypotheses derived from the literature and the research aims. This is, however, not necessarily required according to the guidelines. If this a fully data-driven project, I would suggest to include a sentence stating so and give the rationale, why no hypotheses were given.	Thank you for this remark. We added the hypotheses formulated in the table on pages 21-24 to section 1.**4 Research aims and hypotheses**.
d) 1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable) This generally is well written and follows the standards in the field (PRISMA, etc.).	
Just a few suggestions:	
e) Screening: I would include some statistical measure of inter-rater coherences like Cohen's Kappa.	Thank you for this important suggestion. We will calculate Cohen's Kappa and have indicated this in section **2.5. Screening procedure**.
f) One thing the authors may wish to consider, but is no must: It becomes more and more standard to include formal risk of bias analyses in meta-analyses, e.g. following NOS: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp	Thank you for this important addition. Unfortunately, the proposed NOS are more relevant for clinical intervention studies and hence not quite applicable to our study. However, we agree that a systematic quality assessment is essential and would hence like to adapt the RoBANS 2 scale (Seo et al., 2023). Moreover, we hope to adequately describe study quality within the systematic review and have added a more concrete list of relevant parameters at the end of paragraph 1 in section **2.2 Design**. Finally, we propose to adapt a similar approach to Khalil et al. (2022) which is more catered to DTI data, including more specific markers of study quality in the expected type of studies. Hyun-Ju Seo et al., "RoBANS 2: A Revised Risk of Bias Assessment Tool for Nonrandomized Studies of Interventions," *Korean Journal of Family Medicine* 44, no. 5 (July 7, 2023): 249–60, https://doi.org/10.4082/kjfm.23.0034.

	Khalil, M., Hollander, P., Raucher-Chéné, D., Lepage, M., & Lavigne, K. M. (2022). Structural brain correlates of cognitive function in schizophrenia: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *132*, 37–49. https://doi.org/10.1016/j.neubiorev.2021.11.034
g) 1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses Yes, I think so.	
h) 1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s). I think this is not likely to be an issue in this project.	
i) Evaluation: All together this is a very well-written Stage 1 RR that follows the standards for meta-analyses very well. I think it deserves IPA.	Thank you very much for this very positive feedback and the precise recommendations!
j) Signed, Sebastian Ocklenburg	

3. Reviewer 2: Katie Lavigne, Ph.D.

Reviewer comments	Author response
This is a registered report on a meta-analysis of structural connectivity and social cognition. The authors propose a series of meta-analyses to: (RA1) examine the relationship between social cognition and structural connectivity; (RA2) identify white matter regions associated with social cognition; and (RA3) investigate potential moderators including socio-cognitive constructs (i.e., subdomains), diffusion metrics, and population/diagnosis-specific effects (RA3). Meta-analysis 1 (MA1) will include correlations between diffusion metrics and social cognitive scores (RA1). MA2 will involve a coordinate-based meta-analysis using seed-based d mapping (SDM) from voxel-based and tract-based studies (RA2). MA3 will include correlations between tract-based diffusion metrics and social cognitive measures (RA2). All MAs will be followed by a meta-regression (sex ratio) and subgroup analyses (socio-cognitive construct, diffusion metric, population/diagnosis, age group, whole brain vs. region of interest) to assess RA3. They will also include tests for publication bias and heterogeneity. This study proposes a novel meta-analysis and provides good justification for examining structural connectivity and social cognition. They cite a previous systematic review (Wang et al., 2018), which supports the existence of relevant literature for the proposed meta-analysis. The inclusion of potential moderators is important given the breadth of the proposed meta-analyses, which are expected to have high heterogeneity.	
Major comments:	
k) Has the search strategy been reviewed by an academic librarian? Some terms may be	Thank you for this valuable suggestion. The selected search terms were based on the

too general and/or capture irrelevant areas of research (e.g., social skills/functioning refer more to outcomes than social cognition). Use of wildcards (e.g., social cogniti*) would be appropriate. A preliminary search would ensure feasibility and potentially help revise the search strategy, as too many hits could hinder the screening process.	search terms used by the cited review by Wang et al. (2018) and a meta-analysis on socio-cognitive (Schurz et al., 2021) as well as meta-analyses on structural connectivity. The use of wildcards was now adapted, and the search string has been piloted and revised based on workshop materials provided by the university library.
l) The introduction should include a deeper elaboration of socio-cognitive constructs based on the literature or justification on those selected. It currently includes some examples, such as emotion recognition, theory of mind, and empathy, but omits other areas, such as social perception, social knowledge, and attributional style). This could help guide the search terms for a more comprehensive investigation of social cognition.	Thank you for this important conceptual and practical input. We conceptualize social cognition quite broadly as those processes necessary for social interaction (see Happé et al., 2017), but going beyond simple signal processing and imitation. Upon revision we do agree that more terms would fit into our concept, which has led to the extension of the search string (e.g. by social perception, social motivation, social learning, social knowledge). Additionally, the introduction has been adapted.
m) How will the socio-cognitive measures be categorized into constructs – will this be based on the selected papers (if reported) or done separately by the research team? If the latter, the constructs should be categorized by 2+ experts in social cognition based on the measures/scores used and done separately from the data extraction. The coding sheet should therefore include the score used for the social cognitive measure assessed, to ensure that categorization of social cognitive construct is precise.	Thank you for stressing this important point. As described in section **2.6 Coding**, the socio-cognitive constructs will primarily be recorded as the socio-cognitive measure/assessment tool used (what you refer to as score). This will be entered in the coding template "measure_soc_cog". The column "SoC_construct" will be used to record the constructs discussed/aimed to be assessed by the primary literature. Finally, the resulting constructs will be discussed by the research team (3+ researchers) which includes experts on social cognition. Unfortunately, no standardized list or nomenclature of SoC categories exists so far. Therefore, the experts will consider the constructs proposed by the primary literature as well as work done by authors such as Happé et al. (2017), to group the measures into construct categories.
n) The organization of the series of meta-analyses and follow-up meta-regressions/subgroup analyses could be improved. Although there are 3 research aims and 3 meta-analyses, they do not clearly map onto one another. It appears MA1 will address RA1 and RA3 and that MA2 and MA3 will both address RA2 and RA3. How will the findings regarding RA3 be compared across MAs?	We thank the reviewer for highlighting this important conceptual point. Based on the input within this review round, the research aims were additionally translated into more concrete hypotheses (see review points 2c and 1o). The revised table at the end of the document shall provide an overview of what research aims and questions correspond to which parts of the analyses. Regarding the moderator analysis in the realm of **RA3**, the idea is to differentially assess moderation effects in the different analyses and provide an overview of the results. Since the more fine-grained sub-sample analyses will

	investigate different groups and the moderation effects independently, no integration of these effects is planned. See discussion of multiple comparison below.
o) MA2 involves a seed-based d mapping procedure using voxel and tract-based correlations between diffusion metrics and social cognition. Is this typically reported in the literature at such a fine-grained level? A few citations of relevant papers would show feasibility here.	As described in section 2.7.3, the seed-based d mapping meta-analysis method will be applied. This method has been successfully applied to integrate TBSS and VBA data. Examples would include: Guo et al. (2024) and Yang et al. (2021) Yunxiao Guo et al., "Intrinsic Disruption of White Matter Microarchitecture in Major Depressive Disorder: A Voxel-Based Meta Analysis of Diffusion Tensor Imaging," *Journal of Affective Disorders* 363 (October 15, 2024): 161–73, https://doi.org/10.1016/j.jad.2024.07.050; Chengmin Yang et al., "Microstructural Abnormalities of White Matter Across Tourette Syndrome: A Voxel-Based Meta-Analysis of Fractional Anisotropy," *Frontiers in Neurology* 12 (September 9, 2021), https://doi.org/10.3389/fneur.2021.659250.
p) MA2 also includes mention of downsampling the voxel-based data to TBSS templates; how will this be done without the raw data? Will this reduce MA2 to a tract-based analysis rather than whole brain and, if so, what is the additional value of MA3?	Thank you for this excellent question. We are happy to provide more details on this methodologically rather complex matter: TBSS and VBA approaches both investigate whole brain data. However, TBSS-based analysis uses brain masks that only include what is identified as white matter tracts based on FA maps. This should help to account for prevalent individual variability in the shape of WM tracts. In contrast, VBA uses standard atlases of the whole brain for sample integration. Therefore, the VBA results need to be "downsampled" to TBSS masks to allow for the integration of both whole brain analysis types. The used TBSS template provided by the SDM authors is based on a TBSS skeleton which is a 3D map of the main white matter tracts. It was created by averaging the DTI images of 58 subjects to get the areas of highest FA. As a result of applying this skeleton mask, the analysis of VBA studies is restricted to the white matter tracts that fall within the same areas used by TBSS studies. We would kindly refer you to the cited references which provide rich methodological detail on the feasibility of this approach (Radua et al., 2014; Wise et al., 2016):

	Toby Wise et al., "Voxel-Based Meta-Analytical Evidence of Structural Disconnectivity in Major Depression and Bipolar Disorder," *Biological Psychiatry* 79, no. 4 (February 15, 2016): 293–302, https://doi.org/10.1016/j.biopsych.2015.03.004. (supplementary methods) Joaquim Radua et al., "Anisotropic Kernels for Coordinate-Based Meta-Analyses of Neuroimaging Studies," *Frontiers in Psychiatry* 5 (February 10, 2014): 13, https://doi.org/10.3389/fpsyt.2014.00013. MA3 are effect-size meta-analyses of studies investigating the same WM tract in an ROI-based approach. ROI-based results cannot be integrated sensibly into coordinate-based analyses and will therefore be performed separately. See the following reference for a similar approach: (Yu et al., 2017) Junhong Yu, Charlene L. M. Lam, and Tatia M. C. Lee, "White Matter Microstructural Abnormalities in Amnestic Mild Cognitive Impairment: A Meta-Analysis of Whole-Brain and ROI-Based Studies," *Neuroscience and Biobehavioral Reviews* 83 (December 2017): 405–16, https://doi.org/10.1016/j.neubiorev.2017.10.026. We hope this explanation makes our approach and reasoning clearer and are happy to provide more references in case further detail and information are required.
q) It was difficult to get a sense of how many total meta-analyses would be performed, but it appears as though there may be many. The authors should consider correcting for multiple comparisons given a potentially large number of primary and subgroup analyses.	We appreciate this critical input. Indeed, depending on the number of available studies, our approach could potentially lead to a large number of meta-analyses. This does raise the question of whether multiple testing should be corrected for. However, the more fine-grained analyses investigate different hypotheses since the effects are evaluated in subsamples and construct-, tract- or metric-specific (moderation) effects are investigated within subsamples. The results will only allow for conclusions on the specific subsample (e.g. only FA studies). Therefore, not more than one statistical test is performed for the same but only for slightly different hypotheses (in different samples). The confirmatory hypothesis of whether there is a moderation effect by the

	categories will be analyzed with the sub-group analysis in the global MA1 which includes all studies. The moderation effects in on level MA3 are more exploratory and will serve to generate hypotheses on subsample-differences. For more details on our reasoning we kindly refer to e.g. Bender & Lange (2001) or García-Pérez (2023) Ralf Bender and Stefan Lange, "Adjusting for Multiple Testing—When and How?," *Journal of Clinical Epidemiology* 54, no. 4 (April 1, 2001): 343–49, https://doi.org/10.1016/S0895-4356(00)00314-0. or Miguel A. García-Pérez, "Use and Misuse of Corrections for Multiple Testing," *Methods in Psychology* 8 (November 1, 2023): 100120, https://doi.org/10.1016/j.metip.2023.100120. Especially since a Bayesian approach is used for the majority of the analyses, an alpha correction would not be applicable, and correction is difficult to implement since association between different tests would need to be quantified a priori. See e.g. (Sjölander & Vansteelandt, 2019) Arvid Sjölander and Stijn Vansteelandt, "Frequentist versus Bayesian Approaches to Multiple Testing," *European Journal of Epidemiology* 34, no. 9 (September 1, 2019): 809–21, https://doi.org/10.1007/s10654-019-00517-2. To conclude, we understand your point and share the concerns. However, since the different MAs also investigate distinct hypotheses, either conceptually or in different samples, we would refrain from implementing a correction. We commit to reporting this as a limitation of our design and will provide an overview of all results (not only significant ones) in the Stage 2 report.
Minor comments:	
r) How will articles that use complex statistical techniques be treated, especially ones that involve correlations between brain and behavioural measures (e.g., partial least squares)?	In principle, only correlations between brain and behavioral measures are included. To the best of our knowledge, all effect sizes should be convertible to r2 or Cohen's d scores which should then be integratable.
s) How will articles that include high-risk groups be treated (e.g., relatives of	For all studies, the sample characteristics will be recorded in the population variable.

patients, subclinical treatment-seeking individuals)?	If a study were to investigate such a specific group, this will be documented in the coding sheet in the population/diagnosis column as well. Such specific cases would need to be discussed on a case-to-case basis. If enough studies (>5) were to investigate the same group, this would be a separate category. Such specific populations investigated by less than 5 studies will be treated as "other" and the results for this moderator level will not be discussed in further detail, since the heterogeneity of the group would not lead to sensible results.
t) Articles will be excluded if they fail to report "relevant details on the defined moderators". Is this on any or all of the moderators?	The moderators of interest represent fundamental characteristics of a study (age, sex-ratio, population, DTI-metric, socio-cognitive measure). It will be used as a proxy for study quality if authors fail to report these essential details and therefore, the articles will be excluded.
u) What is the justification for the age groups selected (<20, 20-55, >55)?	We agree that any precise age threshold will retain a certain degree of arbitrariness. However, based upon excellent research such as Bethlehem et al. (2022) and discussion within our team we converged on a definition of age groups that would be coarse enough to allow for a sensible number of studies on each level but still differentiate between important neurodevelopmental stages. Based on literature on brain development (e.g. Arain et al., 2013; Bethlehem et al., 2022), age 20 was chosen as a cut-off for youth because several maturation processes are believed to level off implying more structural stability. Age 55 was chosen as the age where active myelination starts to decrease, structural degradation might onset and early stages of dementia can occur (e.g. Sherin & Bartzokis, 2011). In sum, the chosen age rages are based upon literature on brain maturation as well as discussion within the expert team and considering the expected populations.
v) Given the objective to identify "diagnosis-specific effects", the coding sheet should include the diagnostic category of the sample, in addition to the healthy/patient comparison.	Thank you for this remark. We understand that the phrasing of the moderator "diagnosis or healthy" might have been misleading. As can be seen in the coding sheet, the nominal (not binary) variable *population/diagnosis* will be used to record the investigated diagnoses or other sample characteristic such as *healthy*. This shall allow for a differentiation between different diagnoses as well as healthy.

	We have clarified this in the list of moderators in section **1.4. Research aims and hypotheses** as well as in the model specifications.
w) To guide future research, I would strongly recommend distinguishing between regions/tracts that are non-significant from those that were not assessed when reporting findings. An example can be found in our similar meta-analysis on neurocognition and brain structure (Figure 3): https://doi.org/10.1016/j.neubiorev.2021.11.034	Thank you for this remark and the reference. A qualitative overview of the research landscape will be given in the realm of the systematic review, which will include an overview of the investigated and as a result also the not investigated tracts. We hope our approach satisfies your standards and meets to your concerns.
x) Pre-registration plan should be detailed (PROSPERO? OSF?).	All study material is already available on OSF and will be published as soon as the review of the Stage 1 report is completed.
Signed, Katie Lavigne, Ph.D. Assistant Professor, Department of Psychiatry, McGill University, Montreal, Quebec, Canada Researcher, Douglas Research Centre, Montreal, Quebec, Canada Lead, Douglas Open Science Program	

References specific to review:

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., Sandhu, R., & Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment*, *9*, 449–461. https://doi.org/10.2147/NDT.S39776

Bartoš, F., Maier, M., Stanley, T. D., & Wagenmakers, E.-J. (2023). *Robust Bayesian Meta-Regression—Model-Averaged Moderation Analysis in the Presence of Publication Bias*. OSF. https://doi.org/10.31234/osf.io/98xb5

Bender, R., & Lange, S. (2001). Adjusting for multiple testing—When and how? *Journal of Clinical Epidemiology*, *54*(4), 343–349. https://doi.org/10.1016/S0895-4356(00)00314-0

Bethlehem, R. a. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., … Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, *604*(7906), 525–533. https://doi.org/10.1038/s41586-022-04554-y

Chandio, B. Q., Risacher, S. L., Pestilli, F., Bullock, D., Yeh, F.-C., Koudoro, S., Rokem, A., Harezlak, J., & Garyfallidis, E. (2020). Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations. *Scientific Reports*, *10*(1), 17149. https://doi.org/10.1038/s41598-020-74054-4

Dhollander, T., Clemente, A., Singh, M., Boonstra, F., Civier, O., Duque, J. D., Egorova, N., Enticott, P., Fuelscher, I., Gajamange, S., Genc, S., Gottlieb, E., Hyde, C., Imms, P., Kelly, C., Kirkovski, M., Kolbe, S., Liang, X., Malhotra, A., … Caeyenberghs, K. (2021). Fixel-based Analysis of Diffusion MRI: Methods, Applications, Challenges and Opportunities. *NeuroImage*, *241*, 118417. https://doi.org/10.1016/j.neuroimage.2021.118417

García-Pérez, M. A. (2023). Use and misuse of corrections for multiple testing. *Methods in Psychology*, *8*, 100120. https://doi.org/10.1016/j.metip.2023.100120

Guo, Y., Liu, Y., Zhang, T., Ruan, J., Liu, S., & Ren, Z. (2024). Intrinsic disruption of white matter microarchitecture in major depressive disorder: A voxel-based meta analysis of diffusion tensor imaging. *Journal of Affective Disorders*, *363*, 161–173. https://doi.org/10.1016/j.jad.2024.07.050

Happé, F., Cook, J. L., & Bird, G. (2017). The Structure of Social Cognition: In(ter)dependence of Sociocognitive Processes. *Annual Review of Psychology*, *68*(Volume 68, 2017), 243–267. https://doi.org/10.1146/annurev-psych-010416-044046

Höfler, M. (2021). *Bayesian regions of evidence (for normal distributions)*. OSF. https://doi.org/10.31234/osf.io/mg23h

Maier, M., Bartoš, F., & Wagenmakers, E.-J. (2023). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, *28*(1), 107–122. https://doi.org/10.1037/met0000405

O'Donnell, L. J., Daducci, A., Wassermann, D., & Lenglet, C. (2019). Advances in Computational and Statistical Diffusion MRI. *NMR in Biomedicine*, *32*(4), e3805. https://doi.org/10.1002/nbm.3805

Radua, J., Rubia, K., Canales-Rodríguez, E. J., Pomarol-Clotet, E., Fusar-Poli, P., & Mataix-Cols, D. (2014). Anisotropic Kernels for Coordinate-Based Meta-Analyses of Neuroimaging Studies. *Frontiers in Psychiatry*, *5*, 13. https://doi.org/10.3389/fpsyt.2014.00013

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin*, *147*(3), 293–327. https://doi.org/10.1037/bul0000303

Seo, H.-J., Kim, S. Y., Lee, Y. J., & Park, J.-E. (2023). RoBANS 2: A Revised Risk of Bias Assessment Tool for Nonrandomized Studies of Interventions. *Korean Journal of Family Medicine*, *44*(5), 249–260. https://doi.org/10.4082/kjfm.23.0034

Sherin, J. E., & Bartzokis, G. (2011). Human Brain Myelination Trajectories Across the Life Span: Implications for CNS Function and Dysfunction. In E. J. Masoro & S. N. Austad (Eds.), *Handbook of the Biology of Aging (Seventh Edition)* (pp. 333–346). Academic Press. https://doi.org/10.1016/B978-0-12-378638-8.00015-4

Sjölander, A., & Vansteelandt, S. (2019). Frequentist versus Bayesian approaches to multiple testing. *European Journal of Epidemiology*, *34*(9), 809–821. https://doi.org/10.1007/s10654-019-00517-2

Wang, Y., Metoki, A., Alm, K. H., & Olson, I. R. (2018). White matter pathways and social cognition. *NEUROSCIENCE AND BIOBEHAVIORAL REVIEWS*, *90*, 350–370. https://doi.org/10.1016/j.neubiorev.2018.04.015

Wise, T., Radua, J., Nortje, G., Cleare, A. J., Young, A. H., & Arnone, D. (2016). Voxel-Based Meta-Analytical Evidence of Structural Disconnectivity in Major Depression and Bipolar Disorder. *Biological Psychiatry*, *79*(4), 293–302. https://doi.org/10.1016/j.biopsych.2015.03.004

Yang, C., Yao, L., Liu, N., Zhang, W., Tao, B., Cao, H., Gong, Q., & Lui, S. (2021). Microstructural Abnormalities of White Matter Across Tourette Syndrome: A Voxel-Based Meta-Analysis of Fractional Anisotropy. *Frontiers in Neurology*, *12*. https://doi.org/10.3389/fneur.2021.659250

Yu, J., Lam, C. L. M., & Lee, T. M. C. (2017). White matter microstructural abnormalities in amnestic mild cognitive impairment: A meta-analysis of whole-brain and ROI-based studies. *Neuroscience and Biobehavioral Reviews*, *83*, 405–416. https://doi.org/10.1016/j.neubiorev.2017.10.026