**Thanks to the authors for their responses and clarifications. I am not convinced that the measure of RWA as a unitary construct is appropriate for a VBM study. The authors’ responses to my concerns about how authoritarian submission relates to RWA ideology are based on Altemeyer’s (1998) conceptual framework, but empirical evidence is lacking. The authors note that the RWA scale has been found to demonstrate acceptable psychometric properties; however, that does not mean that RWA, as measured by the RWA scale, is a unitary trait or trait-like characteristic suitable for use in a VBM study. The evidence I have seen does not seem to indicate that measuring RWA in this way is “carving nature at its joints”. Please see my responses to individual points below.**

|  |  |  |
| --- | --- | --- |
| **Page** | **Line(s)** | **Comments** |
| **3** | **5** | I’m not clear what ‘hallmark characteristic’ means in this context. Is authoritarian submission a core component of RWA ideology, such that if one doesn’t exhibit authoritarian submission, then one can’t be categorised as holding RWA ideology? Or is it possible (albeit unusual) to hold RWA ideology but not be obedient to authority? I’d recommend re-phrasing this sentence to clarify.  **We thank the reviewer and have made the necessary edit to clarify authoritarian submission as a covarying trait (with the other two traits) of RWA ideology (p.3):**  **“Altemeyer (1998) conceptualised right wing authoritarianism (RWA) as an ideology that can be understood as a cluster of three covarying traits: authoritarian submission, authoritarian aggression and conventionalism. That is, these traits comprise a singular measure of RWA. Authoritarian submission or the tendency to almost unquestioningly obey an authority figure is one such hallmark trait of RWA.”**  **Thank you for clarifying Altemeyer’s (1998) conceptualisation. Empirical evidence is needed as well though. Research has found that authoritarian submission, authoritarian aggression and conventionalism can be differentially associated with ideological beliefs (Reese, 2012), suggesting that these do not covary to the point of singularity. There is also evidence that conventionalism can be de-coupled from authoritarianism (e.g. Passini, 2017; Torres-Vega et al., 2021)**  **Passini, S. (2017). Different Ways of Being Authoritarian: The Distinct Effects of Authoritarian Dimensions on Values and Prejudice. *Political Psychology*, *38*(1), 73–86.** [**https://doi.org/10.1111/pops.12309**](https://doi.org/10.1111/pops.12309)  **Reese, G. (2012). When Authoritarians Protect the Earth—Authoritarian Submission and Proenvironmental Beliefs: A Pilot Study in Germany. *Ecopsychology*, *4*(3), 232–236.** [**https://doi.org/10.1089/eco.2012.0035**](https://doi.org/10.1089/eco.2012.0035)  **Torres-Vega, L. C., Ruiz, J., & Moya, M. (2021). Dangerous Worldview and Perceived Sociopolitical Control: Two Mechanisms to Understand Trust in Authoritarian Political Leaders in Economically Threatening Contexts. *Frontiers in Psychology*, *12*, 623.** [**https://doi.org/10.3389/fpsyg.2021.603116**](https://doi.org/10.3389/fpsyg.2021.603116) |
|  | **26** | “*authoritarian submission, and by extension the RWA ideology*” – I think there is a logical fallacy here. Even if those who hold RWA ideology always exhibit authoritarian submission, it does not mean that those who exhibit authoritarian submission will always hold RWA.  **We thank the reviewer for this point and we would like to further clarify on how the three traits are related to RWA. The original conceptualisation of the RWA scale is such that theoretically, the three traits are subsumed under a unidimensional construct of RWA, not only because they are correlated but they covary with one another. This is likely to be a result of how the items were phrased. Visual inspection of the scale would reveal that a handful of the items are double or triplebarrelled, in that they measure more than one of the covarying traits.**  **To illustrate, one of the items can be divided into their individual traits: “Our country desperately needs a mighty leader (authoritarian submission); who will do what has to be done to destroy (authoritarian aggression); the radical new ways and sinfulness that are ruining us (conventionalism)”. Nevertheless, both the 32-item and 22-item versions of the scale demonstrate acceptable psychometric properties and are the most widely used version of the RWA scales and were therefore utilised in this manuscript. In this way, the logic of authoritarian submission having a biological basis would extend to the**  **RWA ideology, which has also been noted in other published work (e.g. Warner, Tranel & Asp (2016) The Henchman’s Brain**  **Neuropsychological Implications of Authoritarianism and Prejudice).**  **Other research suggests that authoritarian submission may not be strongly related to ideology however (e.g. Vallerga, 2010).**  **There is also evidence in support of the argument that RWA is not a stable personality dimension, but is better conceptualized as latent predisposition that includes three distinct underlying dimensions, which may fluctuate depending on the contextual level of perceived threat to collective security (Winter et al., 2021).**  **As the authors note, the RWA scale conflates authoritarian submission, authoritarian aggression and conventionalism. Conventionalism in the RWA is also operationalised as conservatism. Recent data (Nemet, 2018) fails to support a single construct underlying all RWA items, and instead suggests that attitudes toward individual freedoms (e.g. sexual, religious, vocational, etc.) may be distinct from authoritarianism (Arikan & Sekercioglu, 2019; Costello et al., 2020).**  **Arikan, G., & Sekercioglu, E. (2019). Authoritarian Predispositions and Attitudes Towards Redistribution. *Political Psychology*, *40*(5), 1099–1118. https://doi.org/10.1111/pops.12580**  **Costello, T. H., Bowes, S., Stevens, S. T., Waldman, I., Tasimi, A., & Lilienfeld, S. O. (2020). *Clarifying the Structure and Nature of Left-Wing Authoritarianism*. PsyArXiv. https://doi.org/10.31234/osf.io/3nprq**  **Nemet, J. (2018). *The Relationship Between Right-Wing Authoritarianism & Support for Military Action Among Millennial Voters*. https://academicworks.cuny.edu/jj\_etds/56**  **Vallerga, M. E. (2010). *Pure Authoritarianism: A New Approach to Authoritarianism* [Master of Arts, San Jose State University]. https://doi.org/10.31979/etd.5xnf-haax**  **Winter, T., Jose, P., Riordan, B., Bizumic, B., Ruffman, T., Hunter, J., Hartman, T. K., & Scarf, D. (2021). *Left-wing support of authoritarian submission to protect against societal threat*. PsyArXiv. https://doi.org/10.31234/osf.io/hu9ef** |
| **8** | **16-18** | This argument seems to undermine the rationale for the study. If self-report measures are insufficient or inaccurate measures of RWA and SDO, then how does it help to look at the neuroanatomical correlates of scores on these self-report measures?  **We thank the reviewer for this suggestion and have made the necessary amendment in the introduction (p.8):**  **“Although traditional self-report measurements of RWA and SDO have demonstrated robust reliability and validity across multiple studies, the examination into the neural bases of RWA and SDO can provide more solid evidence for their status as stable individual differences.”**  **Please see my points above regarding evidence that RWA may not be unidimensional and may be a motivational expression of underlying predispositions, rather than a stable trait.** |
| **8** | **27-32** | This is a little unclear. To me, it reads as though the two predictions are opposing, whereas in fact, they are concurrent predictions.  **We thank the reviewer for this highlighting this point and have made the necessary amendment in the introduction (p.8):**  **“We predict that RWA and SDO would involve identical brain regions as they are both system-justifying ideologies that individuals espouse to maintain the hierarchical structure of society. Additionally, these constructs correlate but are nonetheless independent, and would therefore recruit unique brain regions to differentially substantiate these ideologies in terms of antecedents and outcomes as propounded by the DPM model.”**  **This seems to contradict the later sentence, “Our prediction of non-overlapping neuroanatomical regions associated with RWA and SDO suggests an independence of function between these two ideologies at the neural level.”** |

|  |  |  |
| --- | --- | --- |
| **10** | **29-31** | I don’t follow the reasoning here. It could well be the case that SDO covaries with STS and dlPFC activity when viewing faces of different perceived ranks in real life, regardless of how social rank is defined ?  **We thank the reviewer for this clarification and agree that SDO is likely associated with perceived ranks in real life. However, the finding that SDO scores covary with STS and dlPFC is based on a loose operational definition of ‘superiority in a social hierarchy’. In the case of Ligneul et al.’s study, this was defined as ‘competitive skill’ in the task. That is, the authors implicitly defined ‘winner’ as ‘more superior in a social hierarchy’. Because there was no manipulation check of whether participants perceived ‘winners’ as ‘higher in rank’ than another in a social hierarchy, whether SDO scores covary with STS and dlPFC activity due to perceived ranks is debatable.**  **We have also clarified this in the main text (p.10):**  **“It is possible that participants did not perceive any social ranking during the task at all. As there was no manipulation check for this implicit assumption, it is not clear why SDO scores covaried with dlPFC and STS activity. Consequently, the association between SDO and dlPFC and STS regions may not be borne out once this particular task is no longer carried out during the brain scan.”**  **I understand this point now, thanks to the authors for clarifying.** |

|  |  |  |
| --- | --- | --- |
| **11** | **9** | I don’t understand why H4 includes only ACC, and not midcingulate cortex as well?  The description of Cazzato et al’s findings mentions other regions within the “social orienting circuit” – I’m not sure why the hypothesis is specific to the MCC/ACC & insula, rather than including the whole social orienting circuit?  **We thank the reviewer for bringing this clarification to our awareness. Research has made clear distinctions between ACC and MCC and it was a mistake on our part to assume that one could substitute for the other. Given that only the insula was a region that was found to be implicated with SDO in both the Cazzato et al. and Chiao et al. studies, we opted to focus H4 only on the association between SDO and insula. Additionally, the Cazzato et al paper did not make any specific hypothesis of SDO scores covarying with the whole social orienting circuit so we hesitate to make any assumptions. We have also amended this in the main text (p.11):**  **“Therefore, we believe it’s likely that variation in SDO scores will be negatively associated with structural volume of the insula (H4).”**    **We understand that this might cause some misunderstanding so we have now deleted the paragraph in the introduction where we referred to the social orienting circuit.**  **This seems much clearer now.** |
| **11** | **26** | As far as I can tell, there is insufficient grounds to predict that SDO scores will not correlate with vmPFC. As the of Cazzato et al. study used an ROI analysis, and the Chiao et al study reported activation in PFC.  **We thank the reviewer for this point and have made the necessary elaboration in the main text to explicate the double dissociation of SDO and RWA (p.11):**  **“Our prediction of non-overlapping neuroanatomical regions associated with RWA and SDO suggests an independence of function between these two ideologies at the neural level.**  **Note: This seems to contradict the earlier sentence, “We predict that RWA and SDO would involve identical brain regions”**  **Though there is no direct evidence for this double dissociation, some indirect evidence in the literature hints to this possibility. The study by Asp, Ramachandran & Tranel (2012) demonstrated that only damage to vmFPC was significantly associated with higher RWA scores compared to healthy controls. Patients with damage to other neural structures, including those that are involved with emotion, did not show this increase in RWA scores. The etiologies of these non-vmPFC lesions were not overly specified. Notwithstanding, this distinctiveness of RWA scores associated with only vmPFC damage and not other cortical regions implicated in emotional processing leads us to hypothesise that RWA is likely not to associate with the insula. Importantly, this non-vmPFC lesion group excludes patients with specific damage to the amygdala. Thus, the predicted overlapping association of RWA and SDO with the amygdala remains intact. Moreover, only Chiao et al. (2009) has thus far conducted a whole-brain analysis to identify regions that covary with SDO scores during an fMRI task. SDO scores were a significant predictor of frontal areas, namely, inferior, superior and middle frontal gyri activity, in addition to the aforementioned ACC and insula activity when participants engaged in an empathic task. However, after controlling for age and self-reported dispositional empathy, only the ACC and insula were left as regions significantly associated with SDO scores. To our knowledge, no other studies have conducted a wholebrain analysis involving SDO. Comparing with the study by Cazzato et al. (2016), only the insula region consistently covaries with SDO scores across different fMRI tasks. Based on the limited research on this topic, we hypothesise that SDO but not RWA will be associated with the insula and RWA but not SDO will be associated with vmPFC.”** |

|  |  |  |
| --- | --- | --- |
| **17** | **14** | Sex should be controlled for as well as age  **We agree with the reviewer and we will include age, gender as covariates in the model. As per another reviewer’s recommendation, we have also included global brain volume in the form of total intracranial volume as an additional covariate. Please see the detailed description on p.19:**  **‘’In these analyses, we intend to control for total intracranial volume (TIV), age and gender by including them into the regression model as independent “nuisance” variables. TIV is an important variable to account for particularly in ROI-based volumetric measures because such subtle differences in regional brain volume may be confounded by individual differences in overall brain size (O’Brien et al., 2011). We are also controlling for age not only because TIV varies as a function of age (Bartholomeusz et al., 2002), but also because both RWA and SDO have been shown to decrease with age (Altemeyer, 1998; Ruffman et al., 2020). Accounting for age is also necessitated in this study because the analysis will include participants from two different age groups, a young adult sample and a middle-aged adult sample. We would expect both self-report and volumetric brain differences between these two age groups so including age in the regression model will minimise confounds due to age differences. Finally, past research also suggests a gender difference in self-reports of RWA and SDO. In particular, women tend to report higher RWA scores than men (Brandt & Henry, 2012) whereas men tend to report higher SDO scores than women (Pratto et al., 1994). Combined with an overall brain volume differences between men and women (Kaufmann et al., 2001; Ruigrok et al., 2014; Takahashi et al., 2011), we reckoned controlling for gender would facilitate in identifying significant neuroanatomical correlates, as we predict with the age variable. We would like to emphasise that although system-justifying ideologies and regional (and overall) brain volume do seem to vary with age and gender, these are treated as nuisance variables in the main analysis as they do not comprise the main objectives of the study.”**  **Note that intracranial volumes are entered in a distinct section of SPM’s VBM module, rather than as variables in the specification of the linear model.** |

**We thank the reviewer for this insightful suggestion and have implemented a combined ROI and whole brain analysis approach to our study. We have added an additional subsection, “2.9. Whole Brain Analysis Plan” (p.20):**

**“To supplement the a priori ROI analysis, we will also be conducting a whole brain analysis using the DARTEL package in SPM12. As with the ROI analysis, RWA or SDO scores will be used as contrasts to test significance of regressions coefficients from zero value. Similarly, age, gender and TIV will be included as covariates. Significance thresholds will be set at a peak-level threshold of *p* < 0.05 with family-wise error (FWE) correction, and uncorrected voxel-wise level of *p* < 0.001.”**

**As I noted, the research questions (“Is/are there any brain region/s …) are ones that can only be answered by whole-brain analysis, so it seems the wrong way round to say that the whole-brain analysis will supplement the ROI analyses.**