24/01/2022

Dear Prof Chambers,

Neuroanatomical Correlates of System-justifying Ideologies: A Pre-registered Voxel-based Morphometry Study on Right-Wing Authoritarianism and Social Dominance Orientation

Jan Paolo M. Balagtas, Serenella Tolomeo, Bindiya L. Ragunath, Paola Rigo, Marc H. Bornstein & Gianluca Esposito

Two of three reviewers were available to assess the revised Stage 1 submission. As you will see, one of the reviewers is now satisfied (pending further minor revisions), and we are substantially closer to IPA, but the other reviewer remains skeptical about the assumption that right-wing authoritarianism is a sufficiently unitary construct for a VBM study to be scientifically meaningful. The reviewer provides a helpful and well-referenced review that requires careful attention. To achieve Stage 1 IPA, it is vital that this concern is addressed to my and the reviewer's satisfaction.

We are extremely grateful that the reviewers have reviewed our submission for the second time and as such, we address them in a point-by-point fashion below. Overall, we followed suggestions as best as we could to improve the rationale and coherence of our chosen methodology for our study and offered justification in instances where this was not feasible. We believe that incorporating the revisions has further improved our paper. Our response to the reviewer's comments are marked in 'bold" and the new added text to the manuscript is marked in 'blue'.

Reviewer 1

The authors have now provided the missing details regarding the ROI-definition and statistical analysis. I have still one remaining minor issue regarding "1B": Indeed, Baumgartner et al., 2013 analyzed associations of SDO and brain structure (s. also Supplementary Table S3). Apart from this issue, I have no further comments. Admittedly, I am still a bit confused regarding the assumed unidimensionality of RWA and SDO (which justifies the expected association with similar brain structures) on the one hand, and the assumed independence of RWA and SDO in the study sample on the other hand. I am really curious to see the findings here.

We would like to thank the reviewer for bringing the supplementary materials of the Baumgartner et al., 2013 reference to our attention. ~~This was an oversight on our part.~~

We address the point that at least one study has attempted to identify neuroanatomical correlates of SDO in p.10:

"Additionally, another study by Baumgartner et al. (2013) did not find a significant association between subscales of SDO and volume of dorsolateral medial PFC, a region anatomically proximal to the dlPFC (see supplementary materials, S3). It is thus unlikely that SDO will be associated with this cluster of regions."

The reviewer highlights an important issue regarding understandable confusion associated with discussing the alternative perspective that RWA and SDO are essentially measuring the same construct. Therefore, we endeavour to clarify our stance on this issue on p.8:

"To proceed with investigating the neuroanatomical correlates of RWA and SDO with a strong predictive framework, we believe the core theoretical stance we prescribe bears repetition. Though it is unlikely that RWA and SDO measure a unidimensional construct, it is instructive for this investigation to respect them as system-justifying ideologies with the same goal of maintaining existing social hierarchies albeit achieved in different ways. Therefore, we postulate that the status of RWA and SDO as system-justifying ideologies will manifest as an overlap in at least one brain region. At the same time, the DPM model argues that RWA and SDO are derived from different underlying motivations and observable outcomes (i.e. the different ways existing social hierarchies are maintained). In alignment with this model, we predict that RWA and SDO will also correlate with at least one other brain region independent of one another."

Reviewer 2

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) conceptualized 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** **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**

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

| | 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 (Altemeyer, 1998). 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 triple barrelled, 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 utilized 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

We would like to extend our deepest gratitude to the reviewer for taking pains to systematically score the RWA literature. More recent research has been gaining a better understanding that RWA can be interpreted as three separate dimensions. Unfortunately, with the current dataset, it is not possible to tease apart the three dimensions given the nature of the scale administered. That is, the 22-item version of the RWA scale used (Altemeyer, 2006) uses double and triple-barrelled questions that contain multiple dimensions in a single item. As it stands, we cannot easily make the distinction of which item falls under conventionalism, authoritarian submission and authoritarian aggression (e.g. "7. The only way our country can get through the crisis ahead is to get back to our traditional values, put some tough leaders in power, and silence the troublemakers spreading bad ideas", "10. Our country will be destroyed someday if we do not smash the perversions eating away at our moral fiber and traditional beliefs", "19. Our country will be great if we honor the ways of our forefathers, do what the authorities tell us to do, and get rid of the "rotten apples" who are ruining everything", etc.). This is an important limitation of the current study. One way to go around this is to first subject the responses to the RWA scale to factor analysis. However, because the items themselves are double/triple-barrelled, interpreting/labeling the emerging factors may not be feasible.

This is, of course, a direction for future research that is worth exploring, particularly whether or not the three separate subscales map better at a neural level than the superordinate RWA scale.

Notwithstanding, we believe that there is merit in proceeding with the current analysis using RWA as a singular measure. With regards to our hypotheses, the theoretical frameworks we are basing our predictions on (i.e. system justifying ideologies and dual process model) conceptualise RWA as a singular measure that contrasts with SDO so to remain consistent with how RWA has been studied thus far would allow us to directly test whether or not these frameworks of system-justifying ideologies and DPM are valid at a neural level. On a more empirical perspective, we have good reason to believe that RWA can be viewed and measured as a stable unidimensional trait. For one, the DPM literature has demonstrated that longitudinally, RWA has acceptable test-retest reliability across at least a five-month period after accounting for the dangerous worldview, which according to the DPM model drives levels of RWA (Sibley et al., 2007; Asbrock et al., 2010). Moreover, although the 22-item version cannot be subjected to factor analysis, one study did conduct a set of factor analyses of RWA (and SDO) showing that the multidimensional and unidimensional models of RWA demonstrate acceptable fit to response data granted the items themselves were already divided into their respective subscales (Kandler et al., 2016). This gives credence to our current use of RWA as a unidimensional trait.

Asbrock, F., Sibley, C. G., & Duckitt, J. (2010). Right-wing authoritarianism and social dominance orientation and the dimensions of generalized prejudice: A longitudinal test. *European Journal of Personality*, *24*(4), 324-340.
Kandler, C., Bell, E., & Riemann, R. (2016). The structure and sources of right–wing authoritarianism and social dominance orientation. *European Journal of Personality*, *30*(4), 406-420.
Sibley, C. G., Wilson, M. S., & Duckitt, J. (2007). Effects of dangerous and competitive worldviews on right-wing authoritarianism and social dominance orientation over a five-month period. *Political Psychology*, *28*(3), 357-371.

Following this discussion, we make a clear admission to the limitation but also credence to interpret the 22-item version of the RWA scale as a stable unidimensional trait in the manuscript (p.16):

"At this juncture, it is worth noting that the version of the scale used in this study has one main drawback. As the 22-item version of the RWA scale uses double or triple-barrelled questions, it is not feasible to tease apart which of the covariations (i.e. authoritarian submission, authoritarian aggression, conventionalism) are reflected in the participants' response to each item. That is, the 22-item RWA scale can only be viewed as a unidimensional trait with three underlying covariations. This is important to acknowledge given that recent research into the structure of RWA demonstrates that the three dimensions can be viewed as separate subscales (e.g. Passini, 2017; Torres-Vega et al., 2021). In addition, Arikin & Sekercioglu (2019) also argues that the construct of authoritarianism may be better

		conceptualised as a predisposition as opposed to a stable trait. Nevertheless, we have good reason to believe that there is merit in exploring RWA as a stable unidimensional trait. With regards to our hypotheses, the theoretical frameworks we are basing our predictions on (i.e. system justifying ideologies and DPM model) conceptualise RWA as a singular measure that can be contrasted with SDO, and so to follow how RWA has been studied thus far would allow us to directly test whether or not these frameworks are valid at a neural level. The DPM literature has also employed cross-lagged data to show RWA has acceptable temporal reliability across at least a five-month period, after accounting for the dangerous worldview, which moderates RWA as predicted by the DPM model (Sibley et al., 2007; Asbrock et al., 2010). Moreover, one study conducted a set of factor analyses of RWA (and SDO) showing that both multidimensional and unidimensional models of RWA demonstrate acceptable fit to response data granted the items themselves were already divided into their respective subscales (Kandler et al., 2016). This gives credence to the conceptualisation we adopted of RWA as a relatively stable unidimensional trait in this study."
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."
		We thank the reviewer for highlighting this inconsistency. We have made amendments to the latter sentence (see below).

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. Compared 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."

We thank the reviewer for highlighting this inconsistency in the main text. We have rephrased this statement to emphasise that the non-overlapping neuroanatomical regions are in tandem to the overlapping region (p.11):

"Our prediction of non-overlapping neuroanatomical regions associated with RWA and SDO suggests that these two ideologies also recruit separate brain regions that reflect the different underlying beliefs that underlie RWA and SDO as predicted by the DPM model."

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 point!

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.

We thank the reviewer for this insight. The whole brain analysis should take precedence given the research questions posed. As such, we have switched the order of the subsections to reflect this precedence (p. 20):

"*2.8. Whole Brain Analysis*

We will be conducting an exploratory whole brain analysis using the DARTEL package in SPM12. RWA or SDO scores will be used as contrasts to test significance of regressions coefficients from zero value. 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 $p < 0.001$.

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 difference between men and women (Kaufmann et al., 2001; Ruigrok et al., 2014; Takahashi et al., 2011), we reckoned to control 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.

We intend to measure the mean grey matter volume (GMV). We will investigate the association between GMV and scores in the RWA and SDO scales using multiple regression analyses. Each multiple regression analysis will use ordinary least squares models with the GMV as the dependent variable, and RWA (or SDO) score, gender, age and TIV as independent variables.

2.9 ROI Analysis

To supplement the exploratory whole brain analysis, an a priori ROI analysis will also be conducted. The ROI analyses will focus on these predictions: the gray matter volume of the amygdala will be positively associated with both RWA and SDO scores (H2), gray matter volume of the vmPFC will be negatively associated with only RWA (H3) and gray matter volume of the insula will be negatively associated with only SDO (H4). Specifically, our main independent variables of

interest are the two measures of system-justifying ideologies (i.e. RWA and SDO scores). These regions were selected based on previous research implicating these respective regions to RWA, SDO or both. With respect to the amygdala, this is the only hypothesis that is based on a previously conducted neuroanatomical study (Nam et al., 2017). However, we would like to reiterate that neither the RWA nor SDO scales were analysed in the full sample of the study. Instead, a general system justification scale was used (Kay & Jost, 2003). These items tap on similar beliefs consistent with high RWA (e.g. "In general, you find society to be fair") or SDO (e.g. "Society is set up so that people usually get what they deserve"). It is worth noting that Nam et al. (2017) previously did not find a significant correlation between SDO and amygdala volume. However, the absence of a relationship may likely have been due to the relatively small sample size (N = 37) used to analyse this relationship. Equipped with a more well-powered sample, our study (N = 82) presents a more definitive measure of a neuroanatomical correlate not only of SDO but also RWA. Therefore, we reasoned that RWA and SDO are likely to also correlate with amygdala volume (H2), as does the general system justification scale in the study by Nam et al. The link between vmPFC and RWA is also unique among the three hypotheses in that it is the only one based on a set of neuropsychological studies involving neurology patients with a lesion in this brain region (Asp, Manzel, et al., 2012; Asp, Ramachandra & Tranel, 2012; Asp & Tranel, 2013). This points to a specific role of the vmPFC in modulating RWA ideologies, such that damage to this region leads to a manifested change in both RWA scores and outcomes related to RWA such as increased religious fundamentalism (Asp, Ramachandran & Tranel, 2012) and more generally, a magnified susceptibility to misleading information (Asp, Manzel, et al., 2012). We believe it is not that farfetched to suspect that the regional volume in vmPFC correlates with the degree of ascription to RWA (H3). Finally, the insula was identified primarily from an fMRI study by Chiao et al. (2009) that found this region to correlate significantly with SDO scores during a pain perception task. This same region was detected in another study using a different task so we can be confident that these correlations are not simply an idiosyncrasy of a specific task type (Cazzato et al., 2015). However, it is worth mentioning that the insula was implicated in this latter study as part of the 'social orienting circuit' in the brain but was not reported to be directly correlated to SDO scores. Nevertheless, both studies taken at face value did use tasks that tap on the essence of SDO – that is the preference for dominance in terms of observing the pain of others (Chiao et al., 2009) or perceived similarity in others (Cazzato et al., 2015). Therefore, based on the limited literature on this topic, we argue that there is value in the present structural ROI-based analysis and we do expect SDO scores to correlate with both regional volumes of ACC and insula (H4).

For each participant, we will average voxel-wise GMV values for each ROI individually, which then serves as the dependent variable for our main analyses.

We will investigate the association between GMV and scores in the RWA and SDO scales using ROI multiple regression analyses. As the brain areas in our hypotheses are identified with a strong a priori prediction, the threshold of significance was set at $p < 0.05$, with small volume correction for multiple comparisons in the ROIs. To analyse these regression models, we will be using the MarsBaR toolbox (http://marsbar.sourceforge.net/) to anatomically define ROIs for the three pre-specified brain regions (i.e. amygdala, vmPFC and insula) according to our main hypotheses. The GMVs will be extracted from their ROIs using anatomically defined spheres with a radius of 20mm centered at (MNI: x=−36, y=−9, z=−17) for the left amygdala and at (MNI: x=27, y=12, z=−21) for the right amygdala (Nam et al., 2017). We then average the mean volumes from the left and right amygdala. The ROI of the vmPFC will be anatomically defined as a sphere with a radius of 20 mm centered at (MNI: x = 0, y = 40, z = -18; Li et al., 2017). Finally, the ROI of the insula will be anatomically defined as a sphere with a radius of 20 mm centered at (MNI: x = -45, y = 26, z = -6; Chiao et al., 2009; Cazzato et al., 2015). Similarly, age, gender and TIV will be included as covariates."

In closing, we would like to thank both reviewers for their careful reading and all the suggestions for how we can improve our work. We hope the reviewers now find our registered report Stage 1 conclusive, interesting and ready for Stage 2.

Best Regards,

Prof Esposito and colleagues

Professor

Social & Affective Neuroscience Lab Psychology Program - SSS, Nanyang Technological University, Singapore

Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

Affiliative Behaviour and Physiology Lab Department of Psychology and Cognitive Science, University of Trento, Italy