Research Report

Religious belief and cognitive conflict sensitivity: A preregistered fMRI study

Suzanne Hoogeveen [a,*], Lukas Snoek [b] and Michiel van Elk [a]

[a] *University of Amsterdam, Department of Social Psychology, 1001, NK Amsterdam, the Netherlands*
[b] *University of Amsterdam, Department of Brain and Cognition, 1001, NK Amsterdam, the Netherlands*

ABSTRACT

In the current preregistered fMRI study, we investigated the relationship between religiosity and behavioral and neural mechanisms of conflict processing, as a conceptual replication of the study by Inzlicht et al., (2009). Participants ($N = 193$) performed a gender-Stroop task and afterwards completed standardized measures to assess their religiosity. As expected, the task induced cognitive conflict at the behavioral level and at a neural level this was reflected in increased activity in the anterior cingulate cortex (ACC). However, individual differences in religiosity were not related to performance on the Stroop task as measured in accuracy and interference effects, nor to neural markers of response conflict (correct responses vs. errors) or informational conflict (congruent vs. incongruent stimuli). Overall, we obtained moderate to strong evidence in favor of the null hypotheses that religiosity is unrelated to cognitive conflict sensitivity. We discuss the implications for the neuroscience of religion and emphasize the importance of designing studies that more directly implicate religious concepts and behaviors in an ecologically valid manner.

1. Introduction

Everywhere across the world, in all times and cultures we find people who believe in supernatural beings. Religious beliefs seem highly successful in offering explanations for various phenomena, ranging from how the world originated, to why one had to switch jobs and what happens after one dies. Yet these beliefs are difficult - if not impossible - to support with empirical evidence. In fact, believers are often confronted with widely supported contradicting evidence, for instance evolutionary explanations of the origins of life or reductionistic explanations of their religious experiences. And yet, despite these challenges, most religious believers keep up their faith (Pew Research Center, 2012).

Various scholars have suggested that a mechanism of reduced conflict sensitivity, i.e., detecting the incongruency between two potentially conflicting sources of information, may foster the acceptance and maintenance of religious beliefs. For example, dual-process accounts of religion (Risen, 2016), the predictive processing model (van Elk & Aleman, 2017), and the cognitive resource depletion model (Schjoedt

---

et al., 2013) all assume that religiosity is associated with a reduced tendency for analytical thinking and error monitoring.

Where the dual-process model by Risen (2016) assumes a conflict between intuitive and analytical thinking that is resolved by acquiescing to the intuition, the predictive processing model by van Elk and Aleman (2017) assumes a conflict between prior beliefs and sensory input that is resolved by assigning more weight to priors and suppressing the influence of error signals (and hence mitigating the update of prior beliefs). The cognitive resource depletion model applies the notion of reduced error monitoring specifically to collective religious rituals. According to the model, the combination of a charismatic authority, a high arousal context, and a sequence of causally opaque ritualized behaviors creates optimal circumstances to facilitate a preordained (religious) interpretation of events and reduces the likelihood for idiosyncratic (potentially non-religious) interpretations. These subtle differences seem to predominantly reflect 'a tale of different literatures', possibly due to the fact that the frameworks originate from different disciplines; dual-process models were developed in social psychology, predictive processing in (cognitive) neuroscience, and the cognitive resource depletion model stems from anthropological research. Nevertheless, all three accounts converge on the key idea that a process of reduced conflict detection (or correction) makes individuals less prone to note information that seemingly contradicts their religious worldviews and to update their beliefs in the light of new information. This mechanism could potentially underlie the relative immunity of religious beliefs to criticism based on empirical observations (cf. what Van Leeuwen, 2014 calls 'evidential invulnerability').

Notably, the implicit assumption of most theoretical frameworks appears to be that a mechanism of reduced conflict sensitivity makes people more receptive to being religious. However, it could also be that being religious affects people's sensitivity to conflicting information; religious 'training' inoculates believers against contradictions and violations of their worldview. This notion parallels findings from mindfulness meditation research reporting evidence that meditation training increases cognitive control as it teaches practitioners to suppress irrelevant information (Moore & Malinowski, 2009; Teper & Inzlicht, 2012), with meditation experts showing less activation in brain areas implicated in attention and cognitive control (e.g., the anterior cingulate cortex; Brefczynski-Lewis, Lutz, Schaefer, Levinson, & Davidson, 2007). As such, mindfulness meditation may train practitioners to flexibly suppress irrelevant information — resulting in increased cognitive control. A similar process may be at play in religious training, in which people also engage in mental practices to maintain attention (e.g., meditative prayer) and to inhibit irrelevant (e.g., sinful) thoughts. On the other hand, naturalness of religion accounts posit that religious concepts (e.g., mind-body dualism, supernatural agents) are highly intuitive and that it is in fact non-religiosity that requires cognitive effort to suppress or reject these intuitions (Barrett, 2000; Bloom, 2007; Boyer, 2008; Norenzayan & Gervais, 2013). This implies that 'secular training' (e.g., analytic thinking and scientific reasoning), rather than religious training, involves suppressing intuitive information and

enhancing the salience of analytic alternatives — resulting in increased cognitive control for non-religious compared to religious individuals.

In line with this latter suggestion, several empirical studies found that increased religiosity is related to a decreased cognitive performance, especially when a logically correct response must override a conflicting intuitive response (e.g., in a base-rate fallacy test; Daws & Hampshire, 2017; Good, Inzlicht, & Larson, 2015; Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014; Zmigrod, Rentfrow, Zmigrod, & Robbins, 2019). Other behavioral studies correlated individuals' self-reported level of religiosity with their performance on low-level cognitive control tasks such as the Go/No-go task or the Stroop task. These studies present a mixed bag of evidence; some report a positive relationship (Inzlicht, McGregor, Hirsh, and Nash (2009)), an inconsistent pattern (Inzlicht & Tullett, 2010), or no relationship (Kossowska, Szwed, Wronka, Czarnek, & Wyczesany, 2016) between religiosity and cognitive control (in terms of accuracy and reaction times).

In addition to this behavioral research, a few neuroscientific studies have been conducted on the association between religiosity and conflict sensitivity. For instance, an fMRI study investigated brain responses in devoted religious believers who listened to intercessory prayer. When participants believed that the prayer was pronounced by a charismatic religious authority, they showed a reduced activation of their frontal executive network, including the dorsolateral prefrontal cortex (DLPC) and the ACC, which have been associated with conflict detection (Schjoedt, Stødkilde-Jørgensen, Geertz, Lund, & Roepstorff, 2011). Furthermore, Inzlicht et al. (2009) conducted a series of EEG studies looking at the relation between religiosity and the error-related negativity (ERN; Inzlicht et al., 2009; Inzlicht & Tullett, 2010). Compared to skeptics, religious believers demonstrated a smaller ERN amplitude in response to errors on a color-word Stroop task (Inzlicht et al., 2009). The authors suggest that these findings reflect the palliative effects of religiosity on distress responses: religious believers experience less distress in association with committing an error and this is reflected in a reduced ERN amplitude. There is, however, an open-ended debate on the functional significance of the ERN; while some researchers interpret the ERN primarily as an affective (i.e., distress) signal, others emphasize that it mainly reflects conflict-sensitivity (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Bush, Luu, & Posner, 2000; Carter et al., 1998; Hajcak, Moser, Yeung, & Simons, 2005; Maier & Steinhauser, 2016; Yeung, Botvinick, & Cohen, 2004).

Relatedly, different views have been proposed on how the relation between religiosity and ACC conflict activity should be interpreted; whereas Inzlicht, Tullett, and Good (2011) suggest that ACC activity in this context reflects error distress, Schjoedt and Bulbulia (2011) argue that the interpretation of ACC activity as reflecting purely cognitive conflict sensitivity is more parsimonious. We believe this discussion partly hinges upon the operationalisation of 'conflict'. EEG studies on cognitive conflict have typically studied the ERN as a proxy for ACC activity. The ERN is an *error*-related signal and reflects neural activity associated with incorrect vs. correct responses, i.e., conflict at the level of the behavioral response (hereafter: response conflict). In contrast, fMRI studies on cognitive conflict typically focus on the neural activity

associated with incongruent vs. congruent stimulus trials, i.e., conflict at the level of information processing (hereafter: informational conflict). Although there is often a correlation between response conflict[1] and informational conflict, not all incongruent trials result in errors, nor do all congruent trials by definition result in correct responses. It is therefore important to dissociate between these two levels of conflict and their associated neural activity (cf. Tang, Critchley, Glaser, Dolan, & Butterworth, 2006; van Veen & Carter, 2005).

It thus remains unclear to what extent religiosity is related to a reduced sensitivity for response conflict (e.g., responding with 'green' when it should have been 'red') or to a reduced sensitivity for informational conflict (e.g., seeing the word 'green' printed in a red font). An effect for *response conflict* should be reflected in a relationship between religiosity and the strength of the error–correct Stroop contrast in the fMRI data, which would be a direct replication of the study by Inzlicht et al. (2009) and their proposed framework (Inzlicht et al., 2011; Proulx, Inzlicht, & Harmon-Jones, 2012). An effect for *informational conflict* should be reflected in a relationship between religiosity and the strength of the incongruent–congruent Stroop contrast in the fMRI data. Schjoedt and Bulbulia (2011), for instance, indeed seem to interpret Inzlicht et al.'s results as religious believers' inattention to conflict monitoring. In everyday life, both sources of conflict detection could play a role in the maintenance of religious beliefs, e.g., when a believer simply does not detect the incongruency between different sources of information or when he/she fails to suppress an intuitive but objectively incorrect answer.

Taking the distinction between response conflict and informational conflict into account, here we investigated two different hypotheses regarding the relation between religiosity and cognitive conflict sensitivity: (1) there is a negative relationship between religiosity and ACC activity induced by response conflict (i.e., the incorrect–correct response contrast), and (2) there is a negative relationship between religiosity and ACC activity induced by informational conflict (i.e., the incongruent–congruent Stroop contrast). We note that both hypotheses are not mutually exclusive, as religiosity could be related to both mechanisms of conflict detection.[2]

Although earlier studies provide preliminary evidence for the religiosity–conflict sensitivity relation, we believe the present study –including a conceptual replication of the seminal study by Inzlicht et al. (2009)– is important for the following reasons. First, in order to substantiate the notion that religious believers are characterized by a *general* tendency for reduced conflict sensitivity at the neural level, a significant correlation or inter-group difference should be established. So

far, only three studies found evidence for an inverse relation between religious beliefs and conflict-induced ACC activity; Inzlicht et al. (2009) showed that religious zeal and belief in God were associated with a reduced ERN response and Kossowska et al. (2016) similarly found that religious fundamentalism was related to a reduced N2 response on the Stroop task, albeit only in the uncertainty condition where participants performed the task under undefined time pressure. Another study failed to find a correlation between neurophysiological measures and religiosity (though the authors did find an experimental effect of priming God's forgiving nature on the ERN; Good et al., 2015). Second, with the exception of Good et al. (2015, $n = 108$), all experiments linking religiosity to ACC activity included small samples and were therefore most likely underpowered (i.e., Inzlicht et al., 2009, $n = 28$ [Study 1], $n = 22$ [Study 2]; Kossowska et al., 2016, $n = 37$) Third, the hypothesized relation between religiosity and cognitive conflict is primarily based on either behavioral or EEG data. EEG studies, however, can offer only indirect evidence for the involvement of specific brain areas (Gazzaniga & Ivry, 2013). The use of fMRI may complement the existing findings, as fMRI allows for a higher spatial specificity, and may thus provide more conclusive evidence regarding the role of the ACC in the acceptance and maintenance of religious beliefs. Finally, the current study design allowed us to dissociate between neural effects related to response conflict (i.e., activity predicted by response accuracy) and to informational conflict (i.e., activity predicted by Stroop congruency). This may help to disentangle the 'conflict sensitivity' accounts of religiosity, and hence affords a more precise theoretical interpretation of the existing data.

1.1. Hypotheses

We tested eight hypotheses, four of which were based on our research questions and four that served as 'outcome neutral tests' (Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014). The four outcome neutral tests were used to validate that our task did indeed induce cognitive conflict (reflected in accuracy and Stroop interference effects), that error commission was reflected in ACC activity, and that informational conflict was reflected in ACC activity. The corresponding outcome neutral hypotheses for the behavioral measures were: ($\mathcal{H}_1$) participants are more accurate on congruent compared to incongruent Stroop trials, and ($\mathcal{H}_2$) participants respond faster on congruent compared to incongruent Stroop trials. Outcome neutral hypotheses for the neural measures were: ($\mathcal{H}_3$) errors on the Stroop task induce more ACC activity compared to correct responses, on average across subjects, and ($\mathcal{H}_4$) incongruent Stroop trials induce more ACC activity compared to congruent trials, on average across subjects.

Conditional on establishing the effects related to hypotheses 1–4, we tested four corresponding hypotheses about the relation between religiosity and conflict sensitivity. For the behavioral measures, we hypothesized that ($\mathcal{H}_5$) Stroop accuracy is negatively related to religiosity, and ($\mathcal{H}_6$) Stroop interference (i.e., the difference in RT for incongruent vs. congruent trials) is positively related to religiosity, indicating decreased cognitive performance. We note that, based on the existing literature one could hypothesize both a positive and a

---

[1] Response conflict is here defined as the conflict between the actual and the correct response, rather than the prepotent and the correct response.

[2] Based on the aforementioned theories addressing believers' failure to notice incompatibility between different sources of contradicting information, we would primarily expect a negative association between religiosity and informational conflict (rather than response conflict). However, from an empirical perspective, our study most closely resembles the design by Inzlicht et al. (2009), who measured and obtained support for a relation between religiosity and neural markers of response conflict.

negative relationship between religiosity and conflict detection; on the one hand, religiosity is associated with reduced response conflict and hence smaller interference effects (cf. Inzlicht et al., 2011). On the other hand, religiosity is associated with an increased tendency for intuitive responding, which means that more effort is required to overcome these intuitive response on incongruent Stroop trials, hence larger interference effects should be expected (cf. Pennycook et al., 2014). Despite these divergent theoretical predictions, most studies have not found any association between religiosity and Stroop interference (Inzlicht et al., 2009, Study 1; Inzlicht & Tullett, 2010; Kossowska et al., 2016), except for Study 2 by Inzlicht et al. (2009), in which a positive correlation between religiosity and Stroop interference was reported. Here, in line with the latter finding we hypothesized a positive relationship between religiosity and Stroop interference.

For the neural measures, we hypothesized that ($\mathcal{H}_7$) the size of the error–correct response BOLD signal contrast (i.e., difference in BOLD signal between errors and correct responses) in the ACC is negatively related to religiosity, on average across subjects (cf. Inzlicht et al., 2009), and ($\mathcal{H}_8$) the size of the incongruent–congruent BOLD signal contrast (i.e., difference in BOLD signal between the incongruent and congruent condition) in the ACC is negatively related to religiosity, on average across subjects. All hypotheses were preregistered on the Open Science Framework (see https://osf.io/nspxb/registrations). Finally, we added exploratory whole-brain analyses to explore whether religiosity is associated with conflict-induced neural activity in any other brain areas besides the ACC.

2. Methods

2.1. Reporting

We report how we determined our sample size, all data exclusions (if any), all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.2. Overview

The data for this study had already been collected as part of the Population Imaging (PIoP) project (May 2015 - April 2016), conducted at the Spinoza Center for Neuroimaging at the University of Amsterdam (see Appendix A for a description of the project). An overview of the data collection and analysis procedure is presented in Fig. 1. All hypotheses were formulated independently without any knowledge of the preprocessed data, and the analysis pipeline was developed and preregistered prior to data inspection.[3] The preregistration can be accessed on the OSF (https://osf.io/nspxb/). This folder also contains the script for the Stroop task, the anonymized raw and processed data and the R scripts used to preprocess

the behavioral data and to conduct the confirmatory analyses (including all figures). The preprocessing scripts for the fMRI analysis and the exploratory fMRI analyses can be found at https://github.com/lukassnoek/ReligiosityFMRI. The (uncorrected) brain maps can be found at https://neurovault.org/collections/6139/.

2.3. Participants

Participants were students who were recruited at the University of Amsterdam and received a financial remuneration. Participants were screened for MRI contraindications before MRI data acquisition. The intended number of participants was 250, but due to technical problems during part of the acquisition process, only 244 participants yielded useable MRI data. Of those 244, data from 20 subjects were excluded due to artifacts in the MRI data due to scanner instabilities or errors during export and/or reconstruction of the data. Additionally, 10 participants were excluded because they did not complete the task of interest (i.e., the gender-Stroop task). These exclusions were known at the time of the preregistration.

We entered the analysis phase with data from $N = 214$ participants. Out of these 214, eight participants were excluded –as preregistered– because they did not complete the religiosity questionnaire or lacked data on the covariates of interest (age, gender, and intelligence). We additionally preregistered to exclude participants whose accuracy was lower than 65%, because this indicates performance at chance level. This means that participants who responded correctly on fewer than 63 out of the 96 trials were excluded. Furthermore, participants who did not respond within the response interval on more than 20% of the Stroop trials were also excluded. As the minimum response interval of 4500 ms is assumed to be sufficient for timely responses, missed responses on more than 20% of the trials were taken to indicate that participants did not understand or perform the task adequately. These criteria led to the exclusion of 14 participants, yielding a total sample size of 193. In addition, for the fMRI analyses, there were 21 participants who did not make any mistake during the task, preventing us from calculating the 'incorrect–correct' contrast.[4] As such, the confirmatory ROI and whole-brain analyses of this contrast were based on data from 172 participants. All other analyses were done on a total of $N = 193$ participants with complete data. The final sample consisted of 109 (56.5%) women and 84 (43.5%) men. The average age of the participants was 22.2 years ($SD = 1.9$; range $= 18 - 26$).

The study was approved by the local ethics committee at the Psychology Department of the University of Amsterdam (Project #2015-EXT-4366) and all participants were treated in accordance with the Declaration of Helsinki.

---

[3] Specifically, LS was involved in data collection and (pre)processing the MRI data and has no access to the religiosity data. MvE and SH formulated the research questions and hypotheses without any access to the MRI data.

[4] Of the 21 excluded participants, 19 made no errors and 2 participants made 1 error, but no reliable signal could be extracted for this error trial.
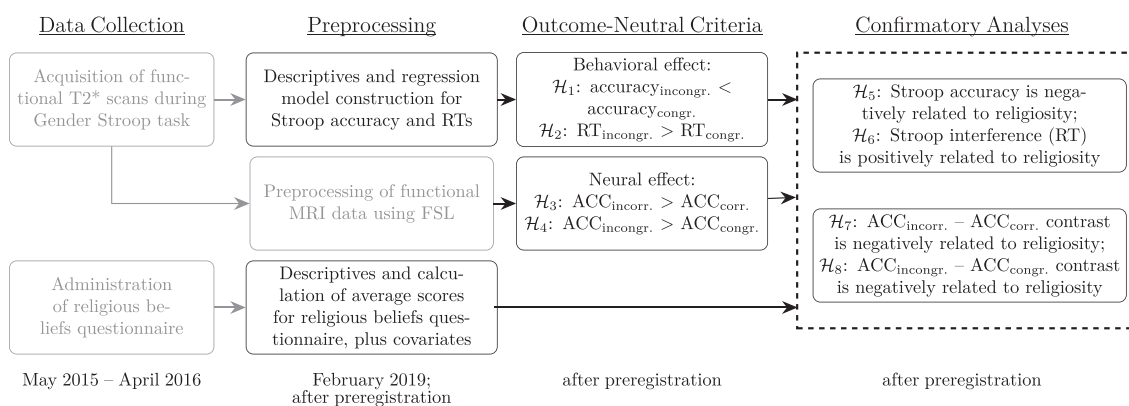
Fig. 1 − **Overview of data acquisition and analysis. Boxes marked in grey had already been completed prior to commencing this project. Boxed marked in black represent the analysis steps for the present study, which were determined in the preregistation.**

2.3.1. Sample size justification

The sample size was determined based on the target of the overall project minus exclusions due to artifacts in the data, incomplete data, or preregistered quality criteria. As there were no existing fMRI studies on the relation between religiosity and cognitive conflict processing —only EEG studies— we could not perform a power analysis. However, we note that a sample of $N \approx 200$ is substantially large for an fMRI study (Szucs & Ioannidis, 2017)[5] and exceeds the recommended minimum sample size of $N = 100$ for correlational (neuroimaging) research (Dubois & Adolphs, 2016; Schönbrodt & Perugini, 2013).

2.4. Procedure

The study ran from May 2015 until April 2016. On each testing day, two participants were tested, which took approximately 4 h and included an extensive behavioral test battery (approximately 2.5 h) and an MRI session (approximately 1.5 h). Participants received a financial remuneration of 50 euros. The order of behavioral and MRI sessions were counterbalanced across participants.

2.5. Study design

The study involved a mixed design with Stroop congruency as the within-subjects variable and religiosity as the between-subjects continuous individual differences variable. The main part of the study qualified as an observational study; we investigated the correlation between performance on the Stroop task and religiosity, and between BOLD-fMRI activity and religiosity, without manipulating any variables except for trial congruency (congruent vs. incongruent Stroop trials). The fMRI task involved a rapid event-related design; a hypothesized BOLD response was modelled following the presentation of facial stimuli in the congruent or incongruent condition, as well as following correct and incorrect responses.
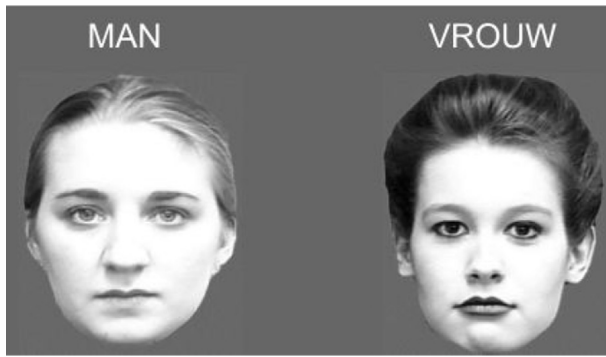
2.6. Stroop task

We used a face-gender variant of the Stroop task (adapted from Egner, Ely, & Grinband, 2010), often referred to as the 'gender-Stroop' task, in which pictures of faces from either gender are paired with the corresponding (i.e., congruent) or opposite (i.e., incongruent) gender label (see below for details on the task and example pictures of the stimuli). The face-gender variant of the Stroop task (Egner & Hirsch, 2005) has been shown to induce significant behavioral conflict and neural ACC activation (Egner, Etkin, Gale, & Hirsch, 2008).[6] Each trial consisted of a photographic stimulus depicting either a male or female face, with the gender label 'MAN' or 'WOMAN' superimposed in red, resulting in gender-congruent and gender-incongruent stimuli (see Fig. 2). The Stroop condition —congruent vs. incongruent— thus formed the within-subjects manipulated variable.

The stimuli set consisted of a total of 12 female and 12 male faces, with the labels 'man', 'sir','woman', and 'lady', both in lower- and uppercase added to the pictures (e.g., 'sir' and 'SIR').[7] All combinations appeared exactly one time, resulting in 96 unique trials (48 congruent and 48 incongruent). Participants were always instructed to respond to the gender of the pictured face, ignoring the distractor word.

---

[5] This meta-analysis reports a median sample size of approximately 22 for fMRI studies.

[6] The face Stroop task - instead of the regular word-color variant - was chosen because it offers optimal opportunities for dissociating between perceptual processing of target and distractor dimensions, as processing of the distractor faces can straightforwardly be linked to activation patterns in the fusiform face area (FFA; Egner & Hirsch, 2005). In the current study, however, we were mainly interested in the cognitive conflict aspect rather than perceptual processing, and therefore solely focused on activation in the ACC.

[7] The Dutch labels were 'man', 'heer','vrouw', and 'dame', respectively.

Fig. 2 – Stimuli as used in the face-gender Stroop task. Distracter words could be incongruent (left) and congruent (right) with the target face. NB. Translations of the Dutch labels: 'MAN' = 'MAN' and 'VROUW' = 'WOMAN'.

The stimuli were presented for 500 ms with a variable inter-trial interval ranging between 4000 and 6000 ms, in steps of 500 ms. Participants could respond from the beginning of the stimulus presentation until the end of the inter-trial interval (i.e., minimum response interval was 4500, maximum response interval was 6500), using their left and right index finger. If no button was pressed during this interval, the trial was recorded as a 'miss'. Stimuli were presented using Presentation (Neurobehavioral Systems, www.neurobs.com), and displayed on a back-projection screen that was viewed by the subjects via a mirror attached to the head coil.

2.7. Religiosity measures

Our religiosity measure consisted of 7 items that were based on religiosity questions included in the World Values Survey (WVS; World Values Survey, 2010), covering religious identification, beliefs, values, and behaviors (institutionalized such as church attendance and private such as prayer). Besides having high face-validity, these measures have been validated in other studies (Lindeman, Svedholm-Hakkinen, & Lipsanen, 2015; Norenzayan, Gervais, & Trzesniewski, 2012; Stavrova, 2015) and the items have been used in previous studies (Maij et al., 2017; van Elk & Snoek, 2020). The items were evaluated on a 5-point Likert scale ranging from *not at all* to *very much*; see Table 1 for the exact items. Ratings on the seven religiosity items were tallied to create an average religiosity score per participant ($M = 1.74$, $SD = 0.84$). Cronbach's alpha for the 7-item religiosity scale was .89, indicating good internal consistency. For the analyses, these average scores were standardized. As anticipated in the preregistration, the distribution of the religiosity data was indeed positively-skewed, since our sample consisted of highly secular students. Although non-normality may reduce statistical power (Poldrack, Mumford, & Nichols, 2011), it does not pose a problem for our analysis, since Bayesian linear regression models —like general (ized) linear models in general— do not assume normality of predictors (solely of model residuals).

Table 1 – Items of the religiosity scale.

1. To what extent do you consider yourself to be religious?
2. To what extent do you believe in God or a supernatural being?
3. To what extent do you believe in life after death?
4. My faith is important to me.
5. My faith affects my thinking and practice in daily life.
6. I pray daily.
7. I visit a church or religious meeting on a weekly basis.

Note. All items were measured on a 5-point scale ranging from *not at all* to *very much*.

2.8. Additional variables

Gender, age, and intelligence were included as covariates in the analyses of the main hypotheses. Intelligence was indexed by the sum score on the 36 item version (set II) of Raven's Advances Progressive Matrices Test (Raven, 2000; Raven, Raven, & Court, 1998). The rationale for including these measures as covariates in our analysis was to control for the potential confound that any religiosity effect may be driven by other individual differences that are known to be associated with religiosity; females are typically more religious than males (Miller & Hoffmann, 1995), older people tend to be more religious than younger people (Argue, Johnson, & White, 1999), and people scoring high on intelligence are on average less religious (Zuckerman, Silberman, & Hall, 2013). Age and intelligence scores were standardized in the analyses.

Since the proposed study was part of a larger project, a number of extra tasks and questionnaires were administered to the participants (see Appendix A for a description). These measures were not included in the present study.

2.9. fMRI data acquisition

Subjects were tested using a Philips Achieva 3T MRI scanner and a 32-channel SENSE headcoil. A survey scan was made for spatial planning of the subsequent scans. After the survey scan, five functional (T2*-weighted BOLD-fMRI) scans (corresponding to five different tasks, including the gender-Stroop task; see Appendix A for an overview of the other tasks), one structural (T1-weighted) scan, and one diffusion-weighted (DWI) scan were acquired. The DWI scan will not be described further, as it is not relevant to the current study. The Stroop task was done during the second scan of the session (not including the survey scan).

The structural T1-weighted scan was acquired using 3D fast field echo (TR: 82 ms, TE: 38 ms, flip angle: 8°, FOV: 240 × 18 mm, 220 slices acquired using single-shot ascending slice order and a voxel size of 1.0 × 1.0 × 1.0 mm). The functional T2*-weighted gradient echo sequences (single shot, echo planar imaging) were run. The following parameters were used for the MRI sequence during the gender-Stroop task: TR = 2000 ms, TE = 27.63 ms, flip angle: 76.1°, FOV: 240 × 240 mm, in-plane resolution 64 × 64, 37 slices (with ascending slice acquisition), slice thickness 3 mm, slice gap .3 mm, voxel size 3 × 3 × 3 mm), covering the entire brain. During the Stroop task, 245 volumes were acquired.

2.10. Preprocessing

Preprocessing was performed using fmriprep version 1.0.15 (Esteban et al., 2019, 2018), a Nipype (Gorgolewski et al., 2011, 2017) based tool. fmriprep was run using the package's Docker interface. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.1.0 (Tustison et al., 2010) and skull-stripped using ants-BrainExtraction.sh v2.1.0 (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer v6.0.1 (Dale, Fischl, & Sereno, 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov, Evans, McKinstry, Almli, & Collins, 2009) was performed through nonlinear registration with the antsRegistration tool of ANTs v2.1.0 (Avants, Epstein, Grossman, & Gee, 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (Zhang, Brady, & Smith, 2001; FSL v5.0.9).

Functional data was motion corrected using mcflirt (Jenkinson, Bannister, Brady, & Smith, 2002; FSL v5.0.9). 'Fieldmap-less' distortion correction was performed by co-registering the functional image to the same-subject T1w image with intensity inverted (Huntenburg, 2014; Wang et al., 2017) constrained with an average fieldmap template (Treiber et al., 2016), implemented with antsRegistration (ANTs). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve & Fischl, 2009) with 9 degrees of freedom, using bbregister (FreeSurfer v6.0.1). Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.1.0) using Lanczos interpolation. Functional data was smoothed with a 5 mm FWHM Gaussian kernel. Many internal operations of fmriprep use Nilearn (Abraham et al., 2014), principally within the BOLD-processing workflow. For more details of the pipeline see http://fmriprep.readthedocs.io.

2.10.1. Quality control

After preprocessing, the MRIQC package (Esteban et al., 2017) was used to generate visual reports of the data and results of several intermediate preprocessing steps. These reports were visually checked for image artifacts, such as ghosting, excessive motion, and reconstruction errors. Participants displaying such issues were excluded from further analysis.

2.10.2. fMRI first-level model

The fMRI timeseries were modelled using a first level (i.e., subject-specific) GLM, using the implementation provided by the nistats Python package (https://nistats.github.io; Abraham et al., 2014; version rel0.0.1b). The GLM included four predictors modelling elements of the task: incongruent trials, congruent trials, correct trials, and incorrect trials. If a participant did not make any mistakes, the 'incorrect trials' predictor was left out. The predictors were convolved with a canonical hemodynamic response function (HRF; Glover, 1999). Onsets for the (in)congruent trial predictors were defined at the onset of the image and had a fixed duration of .5 s. Onsets for the (in)correct trial predictors were defined at the onset of the response. Additionally, six motion regressors (reflecting the translation and rotation parameters in three dimensions) were included as covariates. GLMs were fit with AR1 autocorrelation correction. After fitting the GLMs, the following contrasts were computed: 'incorrect−correct' and 'incongruent−congruent'. The parameters −beta parameters− and associated variance terms from these contrasts were used in subsequent confirmatory ROI analyses and exploratory whole-brain analyses.[8]

2.10.3. fMRI group-level model (exploratory)

In addition to the confirmatory analyses, we also performed an exploratory whole-brain analysis of the effect of religiosity on fMRI activity associated with response conflict (i.e,. $\mathcal{H}_7$) and informational conflict (i.e., $\mathcal{H}_8$). Similar to the confirmatory analyses, in addition to religiosity, the variables age, gender, and intelligence were added as covariates to the model. In the group-level model and in accordance with the 'summary statistics approach', the first-level 'incorrect−correct' and 'incongruent−congruent' contrast estimates represent the dependent variables, while religiosity, age, gender, and intelligence represent the independent variables. For the participants who did not make any error, we could not compute the 'incorrect-correct' contrast and they were thus excluded from the group-analysis of the 'incorrect-correct' contrast.

We used the FSL tool randomise (Winkler, Ridgway, Webster, Smith, & Nichols, 2014) in combination with threshold-free cluster enhancement (Smith & Nichols, 2009) to perform a non-parametric group-analysis of the effect of religiosity. We ran 10,000 permutations. Specifically, we tested for a non-directional (two-tailed) effect of religiosity variable (controlled for the other covariates). In addition, as 'outcome neutral tests', we computed the average of the first-level contrasts ('intercept-only' model) for both the 'incorrect-correct' and 'incongruent-congruent' first-level contrasts. We corrected for multiple comparisons using the distribution of the 'maximum statistic' under the null-hypothesis (i.e., the default in randomise) with a voxel-level α value of .025 (i.e., $\alpha = 0.05$ but corrected for two-sided tests; Chen et al., 2018). We plotted the significant voxels

---

[8] We note that the current design was suboptimal in estimating the effect of informational conflict (but not response conflict) in the fMRI data. Due to insufficient 'jittering' of the interstimulus interval, the first-level predictors for congruent and incongruent trails were strongly negatively corrected ($\bar{r} = -0.9$). While this does not bias our results (the generalized least squares estimator we used is still unbiased), it *does* increase the variance of our first-level results, which in turn reduces the power of finding a correlation of religiosity with the first-level effect of informational conflict (operationalized by the 'incongruent-congruent' contrast). This issue only applies to the 'incongruent-congruent' contrast, not the 'incorrect-correct' contrast (as these predictors are much less correlated with each other, $\bar{r} = -0.2$).

showing either a negative or positive effect of religiosity on a standard MNI152 brain.

2.11. ROI definition

For this study's confirmatory ROI analyses, we used a pre-registered ROI based on a conjunction of a functional ROI, derived from fMRI activity preferentially associated with 'error' (for $\mathcal{H}_3$ and $\mathcal{H}_7$) or 'conflict' (for $\mathcal{H}_4$ and $\mathcal{H}_8$) extracted using Neurosynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011), and an anatomical ROI based on the anatomical coordinates of the ACC, taken from the Harvard–Oxford cortical atlas (Craddock, James, Holtzheimer, Hu, & Mayberg, 2012). The reasons for using a mask based on both a functional and anatomical ROI are twofold. First, the anatomical ROI of the ACC in the Harvard–Oxford atlas (and many others) consists of several putatively functionally different subregions (Gasquoine, 2013; Holroyd et al., 2004; Vogt, 2005). A functional ROI based on the Neurosynth database would resolve this issue of functional ambiguity within a single (anatomical) ROI; however, the Neurosynth maps for 'error' and 'conflict' contain more brain areas than just the ACC (such as the bilateral insula). Therefore, by using the conjunction between the functional ROIs based on Neurosynth and the anatomical ROI of the ACC, we restrict our analyses to a single *anatomical* region that is most likely to be *functionally* relevant for the psychological constructs of interest, i.e., response conflict ("error") and informational conflict ("conflict"). We realize that due to the ambiguity of the term 'conflict' (which may refer to informational conflict or response conflict), the Neurosynth map for 'conflict' will likely also be based on studies involving response conflict. Although not ideal, we believe that this method is the most appropriate way to define our ROI.

Specifically, for our functional ROI, we used the Neurosynth Python package to conduct separate meta-analyses of the terms "error*" and "conflict*", with a frequency threshold of .001[9]. We used the 'association test map' from the meta-analysis output (FDR-thresholded for multiple comparisons at $p < 0.01$), which reflects voxels which are *preferentially* associated with the term 'error' and 'conflict', rather than other psychological constructs. For our anatomical ROI, we used the 'anterior cingulate cortex' region within the Harvard–Oxford cortical atlas. We will define the ACC within this probabilistic atlas as the set of voxels with a nonzero probability of belonging to the ACC. Our final ROI is based on the logical conjunction of these two ROIs (see Fig. 3). For the confirmatory ROI analyses, we averaged the GLM parameters ($\hat{\beta}$, 'beta-values') and associated variance parameters (var$[\hat{\beta}]$) separately for the 'incorrect–correct' ($\mathcal{H}_3$ and $\mathcal{H}_7$) and 'incongruent–congruent' (for $\mathcal{H}_4$ and $\mathcal{H}_8$) first-level contrasts for each participant. These ROI-average parameters were subsequently analyzed in a hierarchical Bayesian regression model (see Statistical Models section for details).
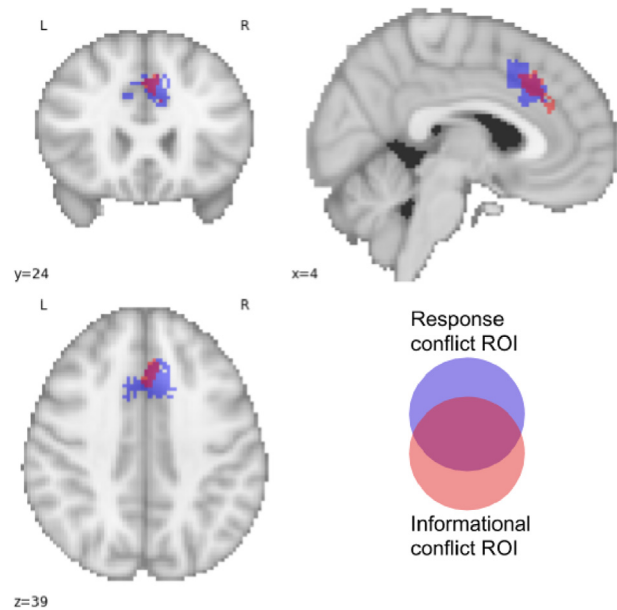
Fig. 3 – ROIs used for our confirmatory ROI analyses of the effect of religiosity on response conflict and informational conflict.

2.12. Statistical models

We applied hierarchical Bayesian models for all hypotheses to accommodate the hierarchical structure of the behavioral and fMRI data, with trials nested within participants. In the multilevel structure, we allow the overall performance and the effect of condition to vary between participants, by including random intercepts and random slopes, respectively. The random intercepts and slopes are desirable theoretically; we are interested in individual differences, hence we should allow effects to differ between individuals. Statistically, omitting the random slope has been shown to result in overestimation of the cross–level interaction term (i.e., the religiosity × condition effect) and the lower level main effect (i.e., the effect of condition; Heisig & Schaeffer, 2019). Finally, adopting this multilevel structure decreases the influence of trial noise through the process of hierarchical shrinkage (see Discussion; Rouder, Kumar, & Haaf, 2019). We constructed the hierarchical Bayesian models using the R package brms (Bürkner, 2017), which relies on the programming language Stan (Carpenter et al., 2017). This package incorporates bridgesampling (Gronau, Singmann, & Wagenmakers, 2017) for hypothesis testing by means of Bayes factors (BF) and posterior probabilities. The general form of our multilevel regression models is:

$$y_{ij} \sim \mathcal{N}\left(\beta_0 + \beta_{0j} + \left(\beta_1 + \beta_{1j}\right)x_{ij}, \sigma^2\right) \tag{1}$$

where y_{ij} is the outcome per trial per participant, and x_{ij} the corresponding value of the predictor. The subscript i is for the individual trials ($i = 1...n_{\text{trials}}$) and the subscript j is for the participants ($j = 1...N$).

2.12.1. Prior specification

We note that the most relevant parameter for making inferences in our specified models is the β_1, i.e., the beta-weight

---

[9] These maps were generated on February 26th, 2019.

for the (standardized) predictors of interest (e.g., Stroop condition, religiosity). As this parameter is used in the critical tests for our hypotheses, it is important to set appropriate priors particularly for this parameter. We chose $\beta_1 \sim \mathcal{N}(0, 1)$ for the (standardized) predictors. This prior is listed as a recommended 'generic weakly informative prior' in the Stan manual (Betancourt, Vehtari, & Gelman, 2015), and has been used in this context before (e.g., Gelman, Lee, & Guo, 2015).

On the remaining parameters we used weakly-informative priors, whereby the priors for the regression weights (β's) are derived from a normal distribution, and the priors on the scale parameters from a half-Cauchy distribution ($\mathcal{C}^+$; Gelman, 2006): $\beta_0 \sim \mathcal{N}(0, 10)$ for the fixed intercept; $\beta_{0j} \sim \mathcal{N}(0, \tau_0^2)$ for the varying part of the intercept per participant; $\beta_{1j} \sim \mathcal{N}(0, \tau_1^2)$ for the varying part of the predictor effect per participant; $\tau \sim \mathcal{C}^+(0, 2)$ for the participant-level variance. Finally, we used the default LKJ-correlation prior to model the covariance matrices in hierarchical models (Lewandowski, Kurowicka, & Joe, 2009). That is, we used $\Omega_k \sim \text{LKJ}(\zeta)$, with Ω_k being the correlation matrix and ζ set to 1.

2.12.2. Interpretation of evidence

Hypothesis testing was done by means of Bayes factors that evaluate the extent to which the data is likely under the alternative hypothesis (e.g., $\mathcal{H}_1 - \mathcal{H}_8$) versus the corresponding null hypothesis $\mathcal{H}_0$. The Bayes factor (BF) reflects the change from prior hypothesis or model probabilities to posterior hypothesis or model probabilities and as such quantifies the evidence that the data provide for $\mathcal{H}_1$ versus $\mathcal{H}_0$, reflected by:

$$\underbrace{\frac{p(\mathcal{M}_1|\text{data})}{p(\mathcal{M}_0|\text{data})}}_{\text{posterior odds}} = \underbrace{\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_0)}}_{\text{prior odds}} \times \underbrace{\frac{p(\text{data}|\mathcal{M}_1)}{p(\text{data}|\mathcal{M}_0)}}_{\text{Bayes factor}} \qquad (2)$$

where $\mathcal{M}_1 - \mathcal{M}_8$ and $\mathcal{M}_0$ represent the models specified for $\mathcal{H}_1 - \mathcal{H}_8$ and $\mathcal{H}_0$, respectively. The Bayes factor BF_{10} then represents the ratio of the marginal likelihoods of the observed data under $\mathcal{M}_1$ and $\mathcal{M}_0$:

$$\text{BF}_{10} = \frac{p(\text{data}|\mathcal{M}_1)}{p(\text{data}|\mathcal{M}_0)} \qquad (3)$$

As our hypotheses are directed, we computed order-restricted Bayes factors, i.e., BF_{+0} in case of an expected positive effect. Note that the subscripts on Bayes factor to refer to the hypotheses being compared, with the first and second subscript referring to the one-sided hypothesis of interest and the null hypothesis, respectively. BF_{+0} is used in case of a hypothesized positive effect for the reference group or a positive relation between variables; BF_{-0} is used for a negative effect for the reference group or a negative relation between variables. As Bayes factors are fundamentally ratios that are transitive in nature, we can easily compute an order restricted Bayes factor; by (1) using the BF for the unrestricted model versus the null model, and (2) comparing the unrestricted model to an order restriction, we can then (3) use the resulting BFs to evaluate the order restriction versus the null model (Morey, 2015).

By default, prior model odds were assumed to be equal for both models that are compared against each other. As the evidence is quantified on a continuous scale, we also present the

results as such. Nevertheless, we included a verbal summary of the results by means of the interpretation categories for Bayes factors proposed by Lee and Wagenmakers (2013, p. 105), based on the original labels specified by Jeffreys (1939). In addition to Bayes factors, we present the posterior model probabilities that are derived from the generated posterior samples.

For all outcome neutral tests we preregistered that a Bayes factor of at least 10 —the minimum value for strong evidence— was required to meet the criteria.

We declare that all models that are described below were constructed before the data were inspected. Additionally, all analyses were run as preregistered. Any deviations are explicitly mentioned in the manuscript.

3. Results

3.1. Outcome neutral tests

3.1.1. Behavioral stroop effect — accuracy

A hierarchical logistic regression model with varying intercepts for the participants and a varying slope for the effect of Stroop congruency was constructed to model response accuracy. In order to validate the presence of a congruency effect on accuracy, i.e., a Stroop effect, we compared the model for $\mathcal{H}_0$ containing only the varying intercept, to the model for $\mathcal{H}-$ containing the varying intercept and the negative effect of congruency. $\mathcal{H}-$ thus indicates that the incongruent condition *decreases* the probability of responding correctly on the Stroop task, relative to the congruent condition.

Results revealed a Bayes factor of 8.43×10^{11} in favor of the alternative model ($\mathcal{M}_-$) relative to the null model ($\mathcal{M}_0$). That is, $\text{BF}_{-0} = 8.43 \times 10^{11}$, indicating that the data are about 10^{11} times more likely under the model assuming lower accuracy for incongruent Stroop trials than for congruent Stroop trials. In order words, the data provide strong evidence for the Stroop effect indexed by accuracy ($\mathcal{H}_1$). See Table 2 for a summary of the results of all four outcome neutral tests.

3.1.2. Behavioral stroop effect — response times

We used a similar hierarchical regression model with varying intercepts for the participants and a varying slope for the effect of Stroop condition to model reaction times. Note that only correct trials are included in the RT analysis. To account for the typical positive skew in RT data, we modelled reaction times as an ex-Gaussian distribution, i.e., a mixture of a Gaussian and an exponential distribution, which has been shown to fit empirical RT data well (Balota & Spieler, 1999; Balota & Yap, 2011; Whelan, 2008). This distribution is incorporated in the brms package, and thus only needed to be specified. Here we expected RTs to be longer for incongruent vs. congruent trials, hence the Bayes factor BF_{+0} was calculated for ratio between the marginal likelihoods of the observed data under $\mathcal{H}_+$ versus $\mathcal{H}_0$. Again, we expected a Bayes factor of at least 10.

We obtained a Bayes factor of 3.53×10^{67} in favor of $\mathcal{M}_+$, that is $\text{BF}_{+0} = 3.53 \times 10^{67}$. In other words, we collected strong evidence for the Stroop interference effect on reaction times ($\mathcal{H}_2$).

Table 2 − Results outcome neutral tests.

Hypothesis	Bayes factor	Posterior Probability	Estimated Coefficient
$\mathcal{H}_1$: accuracy$_{incongr.}$ < accuracy$_{congr.}$	10^{11}	1	− .64 [− .85, − .46]
$\mathcal{H}_2$: RT$_{incongr.}$ > RT$_{congr.}$	10^{67}	1	.03 [.02, .03]
$\mathcal{H}_3$: ACC$_{incorr.}$ > ACC$_{corr.}$	∞[a]	1	3.26 [2.89, 3.64]
$\mathcal{H}_4$: ACC$_{incongr.}$ > ACC$_{congr.}$	157.7	.99	.15 [.03, .26]

Note.

[a] Estimated to approach "infinity" as all posterior samples were in accordance with the order-restricted hypothesis. Bayes factors are the order-restricted Bayes factors for the alternative hypothesis of interest; BF$_{−0}$ for $\mathcal{H}_1$ and BF$_{+0}$ for $\mathcal{H}_2$−$\mathcal{H}_4$. Posterior probabilities are the posterior model probabilities of the alternative model versus the null model. Coefficients are the medians of the posterior distributions for the parameter of interest (i.e., Stroop condition or response accuracy) with 95% credible intervals in square brackets.

3.1.3. Neural processing − response conflict

The hierarchical nature of the fMRI data −being derived from multiple trials− was already taken into account in the calculation of the 'incorrect−correct' contrast and the 'incongruent−congruent' contrast in FSL: we exported the beta-values for each contrast per participant, as well as the variance for the contrasts, i.e., $\widehat{\beta}$ and var$[\widehat{\beta}]$. The inclusion of the variance parameter in the Bayesian models is important, because it allows one to retain the uncertainty associated with the activation level contrast, which is typically lost or ignored when extracting fMRI data for ROI-analyses.[10] In order to test $\mathcal{H}_3$ that the average contrast of ACC activation−the average 'intercept' or $\widehat{\beta}$ − was substantially different from 0, we used the function hypothesis which allows for directed hypothesis test of the specified parameters.[11] $\widehat{\beta}$ is calculated as ($\widehat{\beta}_{incorr.}$ − $\widehat{\beta}_{corr.}$), therefore the hypothesis states that $\widehat{\beta}$ is larger than 0 (i.e., increased ACC activity for errors compared to correct responses). Here we calculated the Bayes factor for $\mathcal{H}_+$ stating that $\widehat{\beta} > 0$.

We note that analyses that took the 'incorrect−correct' fMRI contrast as the dependent variable ($\mathcal{H}_3$ and $\mathcal{H}_7$) include data from 172 participants rather than 193, since some participants made no errors on the Stroop task.

The results showed evidence for the alternative hypothesis to approach "infinity", that is BF$_{+0}$ = ∞. Note that this Bayes factor was estimated by testing the proportion of posterior samples that satisfy the hypothesis that the intercept > 0. When all posterior samples are in accordance with the hypothesis, a Bayes factor of "infinity" can be obtained. In this case that means that the Bayes factor is at least 60,000 since the model included 60,000 samples. In other words, the neural data provide strong evidence that the ACC is sensitive to response accuracy on the Stroop task.

3.1.4. Neural processing − informational conflict

A similar procedure was used to test $\mathcal{H}_4$, this time with the ACC activity contrast for Stroop congruency instead of response outcomes. That is, a hierarchical regression model with a varying intercept for the participants was constructed. The Bayes factor was calculated for the hypothesis that $\widehat{\beta}$ is larger than 0, since we expected $\widehat{\beta}_{incongr.}$ to be larger than $\widehat{\beta}_{congr.}$, resulting in a positive contrast. Again, a Bayes factor of at least 10 was required to pass the outcome neutral criterion test.

A Bayes factor of 157.7 in favor of the alternative hypothesis was obtained (i.e., BF$_{+0}$ = 157.7), indicating that the data provide strong evidence that the ACC is sensitive to informational conflict on the Stroop task.
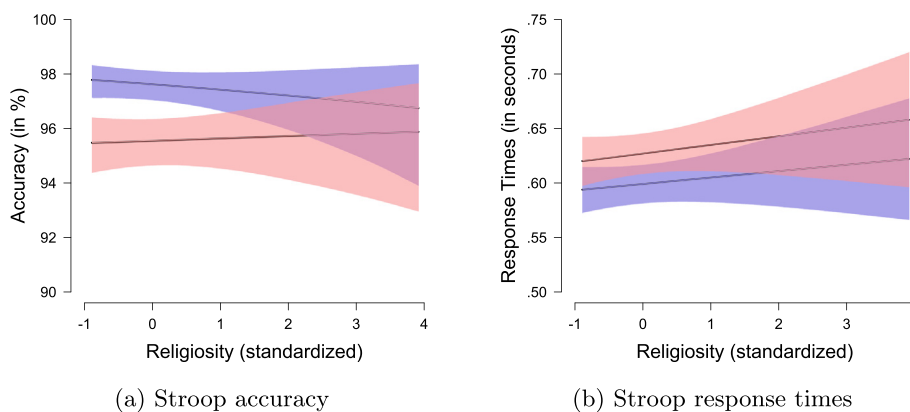
The results of these four analyses indicate that all pre-specified outcome neutral criteria were met.

3.2. Main preregistered analyses

3.2.1. Behavioral stroop effect and religiosity − accuracy

In order to test $\mathcal{H}_5$ whether self-reported religiosity of individuals is related to their performance on a conflict-inducing Stroop task, an extended Bayesian hierarchical logistic regression model was constructed, by adding religiosity as second-level predictor. Specifically, the model for $\mathcal{H}_0$ included varying intercepts and varying slopes for Stroop condition (as before) per participant, plus the participant-level variables gender, age, and intelligence (i.e., the covariates). The model for the alternative hypothesis was identical plus the inclusion of religiosity as an additional participant-level predictor. Notably, an interaction term for religiosity × congruency was also included, as the effect of religiosity might be specific for performance in the conflict condition (i.e., the incongruent Stroop condition). As we expected a negative relation between religiosity and performance on the gender-Stroop task, we restricted the coefficient for religiosity to be negative in calculating the Bayes factor, i.e., we performed a one-sided test.[12] The ratio of marginal likelihoods for the data under $\mathcal{H}−$ versus $\mathcal{H}_0$, i.e., the Bayes factor, were calculated to determine the evidence for the predictive value of religiosity in explaining Stroop performance.

---

[10] The possibility to include the variance of the observations in the regression model formula was added for the purpose of meta-analyses (Vuorre, 2016). However, it also serves the current purpose very well.

[11] The term intercept may be somewhat confusing here. Since the outcome variable is the contrast between the incongruent and congruent condition (i.e., the difference), we only include the intercept in this model, and hence look at the effect of the parameter 'intercept'.

[12] The coefficient for the interaction term was not order-restricted.

(a) Stroop accuracy (b) Stroop response times

Fig. 4 – The marginal effect of religiosity on Stroop accuracy and response time, displayed per Stroop condition. The line with the blue 95% credible interval band indicates performance on congruent Stroop trials, the line with the red 95% credible interval band indicates performance on incongruent Stroop trials.

A Bayes factor of .022 was obtained (i.e., $BF_{-0} = 0.022$, $BF_{0-} = 44.8$), indicating that the data provided more support for the null model than for the religiosity model. This result qualifies as strong evidence that religiosity is not negatively related to accuracy on the Stroop task. The posterior medians and the 95% credible interval for the coefficients of religiosity (-0.08 [$-0.25, 0.09$]) and of religiosity × Stroop condition (0.10 [$-0.04, 0.24$]) indicate that neither religiosity, nor the interaction between religiosity and Stroop condition was related to performance on the Stroop task (see also Fig. 4a). The results of all main hypotheses are also summarized in Table 3. The parameters in the regression models for the four main analyses are displayed in the Appendix (Figure B.7).

3.2.2. Behavioral stroop effect and religiosity – response times

We constructed a similar model with RT as the dependent variable; the model for $\mathcal{H}_0$ was a hierarchical ex-Gaussian regression model for RT with varying intercepts and a varying slope for Stroop condition—including participant gender, age, and intelligence as covariates. For $\mathcal{H}_+$, the model was identical with the added religiosity predictor and the religiosity × congruency interaction term. Again, we hypothesized that religiosity would be negatively related to Stroop performance, hence we expected a *positive* effect of religiosity on Stroop response times.

A Bayes factor of 3.93×10^{-5} was obtained (i.e., $BF_{+0} = 3.93 \times 10^{-5}$, $BF_{0+} = 25461$). Similar to the accuracy analysis, this indicates that the data do not provide support for the hypothesis that religiosity is related to longer response times on the Stroop task. Rather, we obtained strong evidence for the null hypothesis. The posterior medians for the coefficients of religiosity (0.01 [$-0.01, 0.02$]) and of religiosity × Stroop condition (0.00 [$-0.00, 0.01$]) corroborate that there was no main effect of religiosity on response times, nor was there an

interaction of religiosity × Stroop condition on response times (see also Fig. 4b).

3.2.3. Neural processing and religiosity – response conflict

A Bayesian linear regression was performed in order to test $\mathcal{H}_7$ whether self-reported religiosity is related to the ACC sensitivity to incorrect vs. correct responses on the Stroop task. The beta-values for the BOLD contrast in our specified ROI served as the dependent variable, i.e., the extracted $\hat{\beta}$'s. Again, the variance of the individual beta-values was included to take the uncertainty of the contrast estimation into account. Religiosity served as the predictor of interest and gender, age, and intelligence were added as covariates. That is, we compared the model including the contrast-intercept and the covariates ($\mathcal{H}_0$) to the model additionally including the religiosity predictor. Based on the findings by Inzlicht et al. (2009), we expected a *negative* relation between religiosity and ACC activity induced by response conflict.

The results showed more evidence for the null model than for the model including religiosity as a predictor: $BF_{-0} = 0.286$ (i.e., $BF_{0-} = 3.49$). This Bayes factor is interpreted as moderate evidence against the hypothesis that religiosity is associated with reduced ACC sensitivity to response conflict in the Stroop task (i.e., the 'incorrect–correct' contrast). The posterior median and credible interval for the religiosity predictor were -0.09 [$-0.44, 0.26$]. The scatterplot in Fig. 5a illustrates the (absence of an) association between religiosity and sensitivity of the ACC to response conflict.

3.2.4. Neural processing and religiosity – informational conflict

The same model comparison was performed with regard to the stimulus congruency contrast (i.e., $\mathcal{H}_8$). Here, we used the

Table 3 — Results main analyses.

Hypothesis	Bayes factor	Posterior Probability	Estimated Coefficient
$\mathcal{H}_5$: Religiosity ↑ − Stroop performance (accuracy) ↓	.022 (44.82)	.012	− .08 [− .25, .09]
$\mathcal{H}_6$: Religiosity ↑ − Stroop response times ↑	10^{-5} (25461)	.000	.01 [− .01, .02]
$\mathcal{H}_7$: Religiosity ↑ − ACC activity (response conflict) ↓	.286 (3.49)	.172	− .09 [− .44, .26]
$\mathcal{H}_8$: Religiosity ↑ − ACC activity (informational conflict) ↓	.046 (21.87)	.064	.03 [− .09, .15]

Note. Bayes factors are the order-restricted Bayes factors for the alternative hypothesis of interest; BF_{-0} for $\mathcal{H}_5$, $\mathcal{H}_7$, and $\mathcal{H}_8$ and BF_{+0} for $\mathcal{H}_6$. Evidence for the null hypothesis is given between brackets. Posterior probabilities are the posterior model probabilities of the alternative model versus the null model. Coefficients are the medians of the posterior distributions for the parameter of interest (i.e., religiosity) with 95% credible intervals in square brackets.

$\widehat{\beta}$'s of the incongruent−congruent BOLD contrast as the dependent variable. Again, we expected ACC activity to be negatively related to religiosity, while taking into account the effects of gender, age, and intelligence.

A Bayes factor of .046 ($BF_{-0} = 0.046$, $BF_{0-} = 21.9$) was obtained, indicating that the data provide strong evidence against the hypothesis that religiosity is related to reduced ACC sensitivity to informational conflict in the Stroop task (i.e., the 'incongruent−congruent' contrast). The posterior median and credible interval for the religiosity predictor were 0.03 [− 0.09, 0.15]. The scatterplot in Fig. 5b illustrates the (absence of an) association between religiosity and sensitivity of the ACC to informational conflict.

3.3. Exploratory whole-brain analyses

In addition to the confirmatory ROI analyses, we conducted an exploratory (non-parametric) whole-brain analysis of the effect of religiosity on both response conflict and informational conflict. In addition, we ran an 'intercept-only' model (estimating the average effect of response and informational conflict) as an outcome neutral test. All whole-brain *t*-value maps and associated '1-*p*-value' maps can be viewed at and downloaded from Neurovault (https://identifiers.org/neurovault.collection:6139).

3.3.1. Outcome neutral tests
In Fig. 6, we visualized the whole-brain results (as *t*-values) of the 'intercept-only' model for both the response conflict data (i.e., using the 'incorrect−correct' contrast; Fig. 6A) and the informational conflict data (i.e., using the 'incongruent−congruent' contrast; Fig. 6B).

Both whole-brain maps show widespread effects in areas known to be involved in error monitoring and cognitive conflict (such as the ACC and insula). Note that the effects (i.e., *t*-values) are much larger in the response conflict analysis, presumably due to the relatively high variance in the first-level analysis stage due to high predictor correlation.
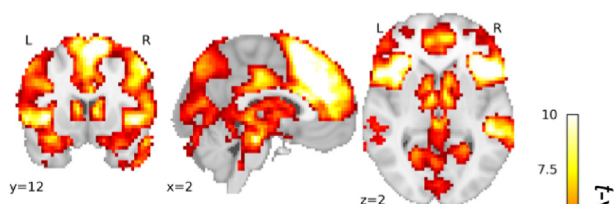
3.3.2. Neural processing and religiosity − response conflict
After multiple comparison correction, no voxels were significantly associated with religiosity in the response conflict analysis.

(a) Religiosity and ACC sensitivity to response conflict

(b) Religiosity and ACC sensitivity to informational conflict
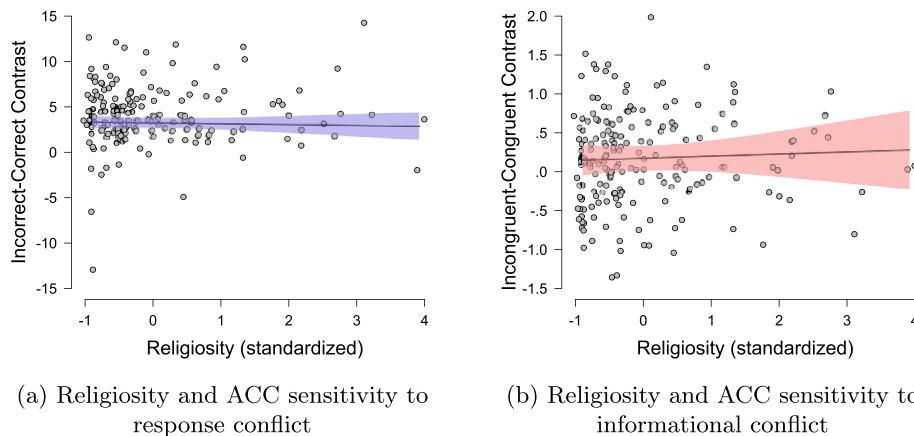
**Fig. 5 — The relation between religiosity on the BOLD signal contrast for incorrect vs. correct responses on the Stroop task (left panel) and on the BOLD signal contrast for incongruent vs. congruent trials in the Stroop task (right panel). The plots display raw individual data points and Bayesian estimated linear effect of religiosity on the conflict-induced BOLD contrasts with 95% credible interval bands.**

A. Response conflict

B. Informational conflict
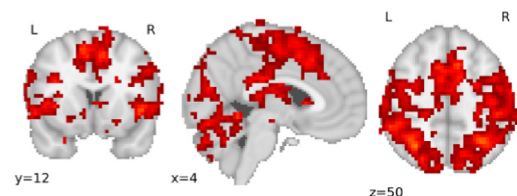
Fig. 6 – Brain maps with *t*-values corresponding to the outcome neutral ('intercept-only') test for both the (A) response conflict analysis and (B) informational conflict analysis. The brain maps were masked using P-values computed using FSL's *randomize* with threshold-free cluster enhancement, which we thresholded at *p* < .05.

3.3.3.   Neural processing and religiosity – informational conflict

Similar to the response conflict analysis, no voxels were significantly associated with religiosity after multiple comparison correction in the informational conflict analysis.

4.   Discussion

In the current preregistered study we investigated whether religiosity is associated with a reduced sensitivity to cognitive conflict as measured through behavioral performance on the Stroop task and neural activation in the anterior cingulate cortex (ACC). The data from the outcome neutral tests provided strong evidence that the gender-Stroop task induced cognitive conflict at the behavioral level ($\mathcal{H}_1$ and $\mathcal{H}_2$) and that this was reflected in increased ACC activity. The neuroimaging data showed that the ACC was responsive to both response conflict (incorrect vs. correct responses; $\mathcal{H}_3$) and informational conflict (incongruent vs. congruent trials; $\mathcal{H}_4$). However, individual differences in religiosity were not related to performance on the Stroop task as measured in accuracy ($\mathcal{H}_5$) and response times ($\mathcal{H}_6$). We also did not observe the hypothesized relation between religiosity and neural activation related to response conflict ($\mathcal{H}_7$) or informational conflict ($\mathcal{H}_8$). Overall, we obtained moderate to strong evidence in favor of the null hypotheses according to which religiosity is unrelated to sensitivity to cognitive conflict. Exploratory whole-brain analyses similarly showed that conflict-induced neural activity was not associated with religiosity.

These results cast doubt on the theoretical claim that religiosity is related to a reduced process of conflict sensitivity. Although this idea is central to various theories about religious beliefs (e.g., Inzlicht & Tullett, 2010; Schjoedt et al., 2013; van Elk & Aleman, 2017), our study shows that religious believers may not be characterized by a *general tendency* of attenuated conflict sensitivity. An important motivation for conducting the current study was to address and overcome the limitations of previous studies in the field. We did so by increasing statistical power (i.e., we used a large sample) and by minimizing degrees of freedom (i.e., we preregistered all hypotheses, methods, and analyses and a priori specified a region of interest (ROI) for the fMRI analysis). Moreover, we curtailed the possibility of (unconscious) biases, as we separated the preprocessing of the fMRI data from the statistical analysis and only combined the fMRI data with the critical variable of interest (i.e., religiosity) in the final analysis steps.

It is important to note that our sample consisted largely of highly secular students; the average religiosity score was 1.74 on a 5-point scale and only 43% considered themselves at least somewhat religious. It could be that the number of religious believers in the sample was simply insufficient to detect an effect. Although this is a serious limitation that nuances the conclusiveness of the current findings, we still believe our study contributes to the existing literature. The fact that the Bayesian analyses showed evidence of absence rather than absence of evidence for the effect, strengthens our belief that previous claims about the association between religiosity and cognitive conflict sensitivity should be interpreted with caution.

Our null findings are perhaps not surprising in light of the recently voiced concerns about the replicability and reliability of neuroscientific findings, often related to problems of insufficiently powered studies (Button et al., 2013; Cremers, Wager, & Yarkoni, 2017; Szucs & Ioannidis, 2017) and general challenges in studying individual differences using neuroimaging (Dubois & Adolphs, 2016). For instance, Boekel et al. (2015) attempted to replicate 17 findings relating behavior to brain structures and found convincing evidence for only one out of the 17 included effects. Similarly, van Elk and Snoek (2020) recently failed to find support for the hypothesized relation between religiosity and grey matter volume in several brain areas that were identified in the literature as being associated with religiosity.

The current study employed the face-gender word variant of the Stroop task rather than the classical color-word Stroop task that has mostly been used in research on religiosity and cognitive conflict sensitivity. Both tasks rely on inhibition of the automatic reading process in order to name the semantic category, with the key distinction that competition takes place either between different features of the same item (i.e., the meaning and the printed color of the word) or between two different items (i.e., the meaning of the word and the 'meaning' of the picture), though also presented within the same visual field. Theoretically, we see no reason to assume that this small difference should be consequential for the religiosity–conflict sensitivity relation; previous claims are

based on a general sensitivity for conflicting information, not exclusively for conflicting features within the same item (as in the color-Stroop task) or in superimposed items (as in the gender-Stroop task). Furthermore, based on the close similarities between the neurocognitive effects associated with both tasks, the picture-word and the color-word Stroop task are often assumed to reflect the same underlying process (e.g., MacLeod, 1991; Starreveld & La Heij, 2017; van Maanen, van Rijn, & Borst, 2009, but see: Dell'Acqua, Job, Peressotti, & Pascali, 2007). Finally, the results of our outcome-neutral tests also provide no indication for substantially different mechanisms at play relative to the classical Stroop task; we find interference effects in the same order of magnitude (i.e., 50.5 ms; Haaf & Rouder, 2019; MacLeod, 1991; Stroop, 1935), and observe the same implicated brain areas (i.e., the ACC, the dorsolateral prefrontal cortex; MacLeod & MacDonald, 2000).

The fact that we did not find behavioral evidence for impaired nor for enhanced Stroop performance among religious believers might indicate that religiosity is unrelated to low-level cognitive control processes. At the same time, the null finding may also reflect the paradox that highly robust experimental effects —such as the Stroop effect— are often difficult to relate to reliable individual differences, irrespective of the specific individual difference construct of interest (Hedge, Powell, & Sumner, 2018; Rouder et al., 2019). That is, because these effects are very robust and automatic ("everybody Stroops"), the between-subjects variability is by definition relatively small. For correlational designs, this 'problem' of small between-subjects variability is further complicated by the presence of measurement error. Rouder et al. (2019) demonstrated that the ratio of true variability (i.e., true differences between individuals) to trial noise (i.e., measurement error) is $1:7$. This unfavorable ratio renders the mission to uncover individual differences in cognitive tasks difficult, if not even impossible. Hierarchical models could mitigate these problems, as these models minimize the effect of trial noise by pulling the trial-level estimates toward the individual's mean effect (known as hierarchical shrinkage). In the current study, we did apply hierarchical modeling for the response time models, as well as the neural ACC models (incorporated in the first-level fMRI models in FSL and by adding the variance parameter of the beta's in the statistical models). Nevertheless, as acknowledged by Rouder et al. (2019), characterizing the degree of measurement error does not imply that the real underlying individual differences can be recovered. This casts doubt on the feasibility to detect true individual variation in cognitive control tasks, and hence to uncover associations with other measures. For example, Hedge et al. (2018) reported correlations of Stroop performance with other measures of cognitive control (e.g., Flanker task, Go/No-go task) ranging from − .14 to .14, none of which were significant. If we cannot even establish correlations between two tasks designed to measure exactly the *same* underlying phenomenon (i.e., cognitive control), the quest for reliable correlations between Stroop performance and more distant constructs such as religiosity seems all the more futile.

Although we obtained moderate to strong evidence for all null hypotheses related to religiosity and cognitive conflict, the current study does not imply that we should reject the notion of reduced conflict sensitivity as a defining characteristic of religious beliefs all together. It could well be that the relationship between religiosity and conflict sensitivity is restricted to specific instances or contexts and hinges strongly on the specific measures and operationalizations that are used. For example, in the study by Good et al. (2015) participants read a sermon about different qualities of God and then performed a Go/No-Go task with alcohol-related stimuli for which responses should be inhibited. As all participants refrained form alcohol consumption in their daily lives based on religious grounds, errors on the Go/No-Go task were seen as 'religious' errors, exposing participants' ostensible pro-alcohol tendencies. The results showed that emphasizing the loving and forgiving nature of God reduced the ERN amplitude in response to religious errors, while emphasizing divine punishment did not affect the ERN compared to a control condition. In other words, it could well be that when participants first contemplate on the comforting nature of their religious beliefs, this may reduce conflict-related ACC activity as induced by a task that includes religion-relevant items and responses. Such a task has much higher ecological validity than the Stroop task that we employed in the current study following the work by Inzlicht et al. (2009). Similarly, the observed reduction of activity in religious believers' DLPC and ACC while listening to a charismatic religious authority (Schjoedt et al., 2011), may specifically depend on the religious content of the speech (and may disappear when the same religious authority would talk about public transport or gardening). It is thus important to do justice to the subjective nature of religious practices and experiences, when studying these topics. This resonates with concerns about the lack of ecological validity in many neuroscience studies on religion (e.g., Schjoedt and van Elk, in press): while studies such as the present one offer high experimental control, the measures do not capture the 'true stuff' that most psychologists and neuroscientists of religion are interested in, namely lived religious beliefs and experiences.

We see two broad future directions for the field. First, the development of new and sophisticated techniques in neuroscience could allow for interesting new hypotheses and measures. For instance, the use of multi-voxel pattern analysis (MVPA) may provide insight into the representational nature of religious concepts endorsed by believers; a question could be whether the neural representations of religious agents such as 'God', 'angels', or 'Satan' are more similar to real people such as 'Napoleon' and 'Donald Trump' or to imaginary agents such as 'Santa Claus' and 'Superman' (cf. Leshinskaya, Contreras, Caramazza, & Mitchell, 2017).

Novel methods for assessing brain connectivity also allow for the investigation of new questions (e.g., Huntenburg, Bazin, & Margulies, 2018; Margulies et al., 2016) One could assess for instance the relationship between

religiosity and the integration of information from sensory cortical areas and the default mode network (DMN), a network that is implicated in abstract, high-level thinking. A hypothesis could be that religious believers are more likely to show a dissociation between the DMN and primary sensory areas. This could be studied in a correlational resting-state design, or alternatively, one could assess believers' brain connectivity while engaging in contemplation of their (religious) beliefs or actions. For instance, intense personal prayer may be associated with a decoupling of internal self-referential processing in the DMN and perceptual processing in the sensory cortices specifically during the prayer experience, similar to what was found for shamanic trance-experiences (Hove et al., 2015).

Second, and relatedly, we believe there is much promise in future endeavours that focus on the application of paradigms and tasks that have higher ecological validity and more closely implicate religious concepts, as in the examples given above. Such an approach can hopefully do more justice to the multifaceted nature of religious beliefs and practices and can pave the way for a truly better understanding of the mechanisms and processes involved in religiosity.

5. Conclusion

In the current study, we attempted to replicate previous findings linking religiosity to a reduced sensitivity to conflicting information, while mitigating limitations of previous work. Our results cast doubt on the general claim that religious believers are characterized by a reduced conflict sensitivity. That is, in contrast to prior research, our data provided evidence *against* an inverse association between religiosity on the one hand and behavioral performance on the Stroop task and induced neural activation in the anterior cingulate cortex (ACC) on the other hand. Although our findings yield no definitive answer to the question of whether and how religiosity may be related to (neural markers of) conflict sensitivity, we believe they highlight that previous claims should be interpreted with caution and that the field could benefit from more ecologically valid measures to investigate this topic.

Funding

CRediT authorship contribution statement

Suzanne Hoogeveen: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing - original draft. **Lukas Snoek:** Data curation, Formal analysis, Methodology, Software, Visualization, Writing - review & editing. **Michiel van Elk:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing - review & editing.

Open practices

The study in this article earned Open Materials, Open Data and Preregistered badges for transparent practices. Materials and data for the study are available at https://neurovault.org/collections/6139/ and https://osf.io/nspxb/.

Appendix A. Population Imaging of Psychology project

The data for this study was collected as part of the Population Imaging of Psychology project (PIoP), which was conducted at the Spinoza Center for Neuroimaging at the University of Amsterdam. The aim of the PIoP was to offer researchers the opportunity to collect brain-imaging data from a large sample of participants (intended N = 250), in association with their individual difference measure of interest. The data was collected between May 2015 and April 2016.
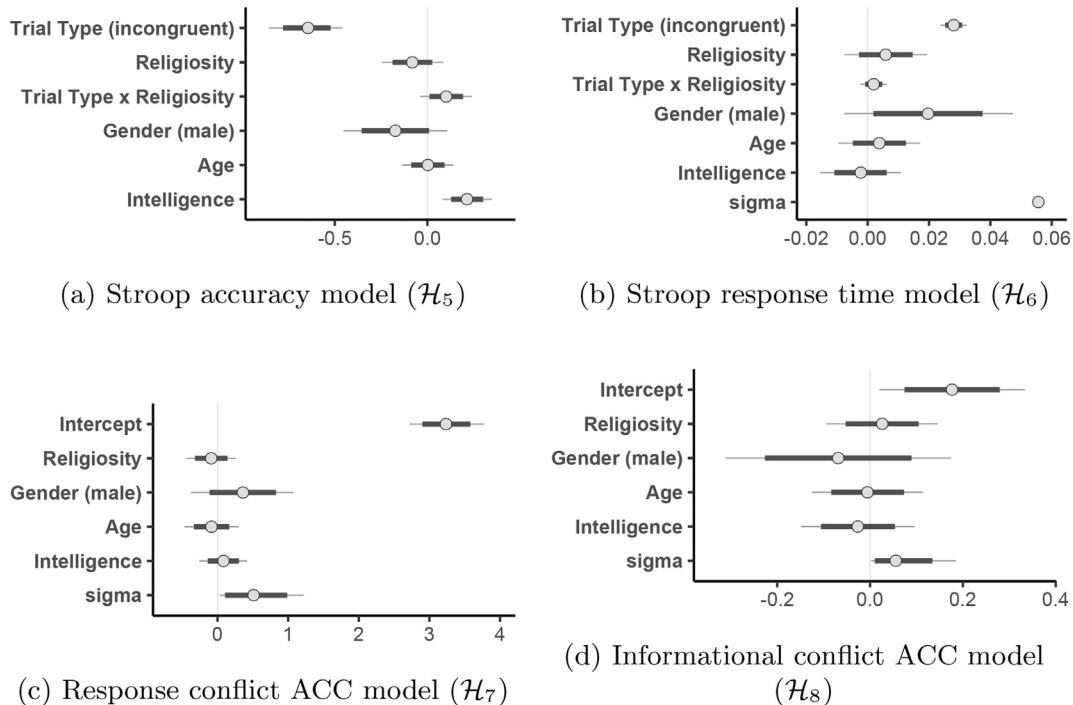
Standard measurements that were collected for each participant included a structural T1 MRI scan, task-free resting state fMRI (6 min), a diffusion tensor imaging (DTI) scan, and different functional localizer scans that were collected using EPI sequences, including the Gender Stroop task, an emotional matching task (Hariri, Bookheimer, & Mazziotta, 2000), a working memory task (Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002), and the anticipation of positively vs. negatively valenced stimuli (Oosterwijk, 2017). In addition, demographic variables were recorded (gender, age, socio-economic status) for each participant, as well as two personality questionnaires, the NEO-FFI (Costa & McCrae, 1992) and the SCID (First, Gibbon, Spitzer, & Benjamin, 1997), and an intelligence test (Raven's matrices; Raven, 2000). Finally, measures on religiosity and religious experiences were included (see Methods for details on the religiosity scale that was used in the present study).

Appendix A.1. Additional Religiosity Items

1. To what extent do you consider yourself to be spiritual?
2. To what extent do you believe in paranormal phenomena (e.g., astrology or telepathy)?
3. To what extent are your parents religious?
4. To what extent do your parents frequently visit a church or religious meeting?
5. Do your parents have a religious lifestyle (e.g., don't go shopping on Sunday, pray before dinner)?

Appendix B. . Coefficient Plots

(a) Stroop accuracy model ($\mathcal{H}_5$)

(b) Stroop response time model ($\mathcal{H}_6$)

(c) Response conflict ACC model ($\mathcal{H}_7$)

(d) Informational conflict ACC model ($\mathcal{H}_8$)

Fig. B.7 – **Coefficients of the fixed effects on Stroop accuracy (top left panel), Stroop response times (top right panel), response conflict ACC activity (bottom left panel), and informational conflict ACC activity (bottom right panel), derived from the Bayesian regression models. For each predictor, points represent the median estimates, thick lines the 80% credible interval and thin lines the 95% credible interval. Note that predictors in the accuracy model are on a linear scale and should be transformed by the inverse logit link to reflect probabilities. In the accuracy and response time models, the intercepts are omitted to enhance visibility of the predictors.**

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics, 8,* 14.

Argue, A., Johnson, D. R., & White, L. K. (1999). Age and religiosity: Evidence from a three-wave panel analysis. *Journal for the Scientific Study of Religion,* 423–435.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis, 12,* 26–41.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128,* 32–55.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20,* 160–166.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences, 4,* 29–34.

Betancourt, M., Vehtari, A., & Gelman, A. (2015). Prior choice recommendations. https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations.

Bloom, P. (2007). Religion is natural. *Developmental Science, 10,* 147–151.

Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, A. J., Brown, S. D., & Forstmann, B. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex, 66,* 115–133.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Evaluating the demand for control: Anterior cingulate cortex and conflict monitoring. *Psychological Review, 108,* 624–652.

Boyer, P. (2008). Religion: Bound to believe? *Nature, 455*(7216), 1038–1039.

Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B., & Davidson, R. J. (2007). Neural correlates of attentional expertise in long-term meditation practitioners. *Proceedings of the National Academy of Sciences, 104,* 11483–11488.

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80,* 1–28.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4,* 215–222.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14,* 365–376.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76.*

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280,* 747–749.

Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience, 1*, 4—17.

Chen, G., Cox, R. W., Glen, D. R., Rajendra, J. K., Reynolds, R. C., & Taylor, P. A. (2018). *A tail of two sides: Artificially doubled false positive rates in neuroimaging due to the sidedness choice with t-tests* (pp. 1037—1043). Human Brain Mapping.

Costa, P. T., & McCrae, R. R. (1992). *NEO personality inventory—revised (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping, 33*, 1914—1928.

Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PloS ONE, 12*, e0184923.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage, 9*, 179—194.

Daws, R. E., & Hampshire, A. (2017). The negative relationship between reasoning and religiosity is underpinned by a bias for intuitive responses specifically when intuition and logic are in conflict. *Frontiers in Psychology, 8*, 2191—2191.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review, 14*(4), 717—722.

Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in Cognitive Sciences, 20*, 425—443.

Egner, T., Ely, S., & Grinband, J. (2010). Going, going, gone: Characterizing the time-course of congruency sequence effects. *Frontiers in Psychology, 1*, 154.

Egner, T., Etkin, A., Gale, S., & Hirsch, J. (2008). Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cerebral Cortex, 18*, 1475—1484.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task—relevant information. *Nature Neuroscience, 8*, 1784—1790.

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). Mriqc: Advancing the automatic prediction of image quality in mri from unseen sites. *PLoS ONE, 12*(9), e0184661.

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., et al. (2018). *poldracklab/fmriprep: 1.0.15*.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019). FMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods, 16*, 111.

First, M. B., Gibbon, M., Spitzer, R. L., & Benjamin, L. S. (1997). *User's guide for the structured clinical interview for DSM-IV Axis II personality disorders: SCID-II*. American Psychiatric Pub.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage, 47*, S102.

Gasquoine, P. G. (2013). Localization of function in anterior cingulate cortex: From psychosurgery to functional neuroimaging. *Neuroscience and Biobehavioral Reviews, 37*, 340—348.

Gazzaniga, M., & Ivry, R. B. (2013). *Cognitive neuroscience: The biology of the mind: Fourth international student edition*. WW Norton.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1*, 515—534.

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics, 40*, 530—543.

Glover, G. H. (1999). Deconvolution of impulse response in event—related BOLD fMRI. *Neuroimage, 9*(4), 416—429.

Good, M., Inzlicht, M., & Larson, M. J. (2015). God will forgive: Reflecting on God's love decreases neurophysiological responses to errors. *Social Cognitive and Affective Neuroscience, 10*, 357—363.

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics, 5*, 13.

Gorgolewski, K. J., Esteban, O., Ellis, D. G., Notter, M. P., Ziegler, E., Johnson, H., et al. (2017). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *13*, 1, 0.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage, 48*, 63—72.

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). *Bridgesampling: Bridge sampling for marginal likelihoods and Bayes factors*. https://github.com/quentingronau/bridgesampling.

Haaf, J. M., & Rouder, J. N. (2019). Some do and some don't? Accounting for variability of individual difference structures. *Psychonomic Bulletin & Review, 26*(3), 772—789.

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology, 42*, 151—160.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *Neuroreport, 11*, 43.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166—1186.

Heisig, J. P., & Schaeffer, M. (2019). Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *European Sociological Review, 35*(2), 258—279.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G., et al. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience, 7*, 497.

Hove, M. J., Stelzer, J., Nierhaus, T., Thiel, S. D., Gundlach, C., Margulies, D. S., et al. (2015). Brain network reconfiguration and perceptual decoupling during an absorptive state of consciousness. *Cerebral Cortex, 26*, 3116—3124.

Huntenburg, J. M. (2014). *Evaluating nonlinear coregistration of BOLD EPI and T1w images*. PhD thesis. Freie Universität Berlin.

Huntenburg, J. M., Bazin, P.-L., & Margulies, D. S. (2018). Large-scale gradients in human cortical organization. *Trends in Cognitive Sciences, 22*, 21—31.

Inzlicht, M., McGregor, I., Hirsh, J. B., & Nash, K. (2009). Neural markers of religious conviction. *Psychological Science, 20*, 385—392.

Inzlicht, M., & Tullett, A. M. (2010). Reflecting on God: Religious primes can reduce neurophysiological response to errors. *Psychological Science, 21*, 1184—1190.

Inzlicht, M., Tullett, A. M., & Good, M. (2011). The need to believe: A neuroscience account of religion as a motivated process. *Religion, Brain & Behavior, 1*, 192—212.

Jeffreys, H. (1939). *Theory of probability* (1 edition). Oxford, UK: Oxford University Press.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*, 825—841.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology, 13*, e1005350.

Kossowska, M., Szwed, P., Wronka, E., Czarnek, G., & Wyczesany, M. (2016). Anxiolytic function of fundamentalist

beliefs: Neurocognitive evidence. *Personality and Individual Differences, 101,* 390–395.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course.* Cambridge (UK): Cambridge University Press.

Leshinskaya, A., Contreras, J. M., Caramazza, A., & Mitchell, J. P. (2017). Neural representations of belief concepts: A representational similarity approach to social semantics. *Cerebral Cortex, 27,* 344–357.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100,* 1989–2001.

Lindeman, M., Svedholm-Hakkinen, A. M., & Lipsanen, J. (2015). Ontological confusions but not mentalizing abilities predict religious belief, paranormal belief, and belief in supernatural purpose. *Cognition, 134,* 63–76.

van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review, 16*(6), 987–999.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383–391.

Maier, M. E., & Steinhauser, M. (2016). Error significance but not error expectancy predicts error-related negativities for different error types. *Behavioural Brain Research, 297,* 259–267.

Maij, D. L. R., van Harreveld, F., Gervais, W., Schrag, Y., Mohr, C., & van Elk, M. (2017). Mentalizing skills do not differentiate believers from non-believers, but credibility enhancing displays do. *PLoS ONE, 12,* e0182764.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences, 113,* 12574–12579.

Miller, A. S., & Hoffmann, J. P. (1995). Risk and religion: An explanation of gender differences in religiosity. *Journal for the Scientific Study of Religion, 34,* 63–75.

Moore, A., & Malinowski, P. (2009). Meditation, mindfulness and cognitive flexibility. *Consciousness and Cognition, 18,* 176–186.

Morey, R. D. (2015). *Multiple comparisons with BayesFactor, Part 2 — order restrictions.* https://richarddmorey.org/category/order-restrictions/.

Norenzayan, A., & Gervais, W. M. (2013). The origins of religious disbelief. *Trends in Cognitive Sciences, 17,* 20–25.

Norenzayan, A., Gervais, W. M., & Trzesniewski, K. H. (2012). Mentalizing deficits constrain belief in a personal God. *PLoS ONE, 7,* e36880.

Oosterwijk, S. (2017). Choosing the negative: A behavioral demonstration of morbid curiosity. *PLoS ONE, 12,* e0178399.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). Cognitive style and religiosity: The role of conflict detection. *Memory & Cognition, 42,* 1–10.

Pessoa, L., Gutierrez, E., Bandettini, P. A., & Ungerleider, L. G. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron, 35,* 975–987.

Pew Research Center. (2012). *The global religious landscape.* https://www.pewforum.org/2012/12/18/global-religious-landscape-exec/.

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional MRI data analysis.* Cambridge University Press.

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences, 16,* 285–291.

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41,* 1–48.

Raven, J., Raven, J. C., & Court, J. H. (1998). Manual for Raven's progressive matricesand vocabulary scales. In *Chapter section 4: The advanced progressive matrices.* Oxford, England: Oxford Psychologists Press.

Risen, J. L. (2016). Believing what we do not believe: Acquiescence to superstitious beliefs and other powerful intuitions. *Psychological Review, 123,* 182–207.

Rouder, J. N., Kumar, A., & Haaf, J. M. (2019). Why most studies of individual differences with inhibition tasks are bound to fail. *Preprint via PsyArXiv.*

Schjoedt, U., & Bulbulia, J. (2011). The need to believe in conflicting propositions. *Religion, Brain & Behavior, 1,* 236–239.

Schjoedt, U., Sørensen, J., Nielbo, K. L., Xygalatas, D., Mitkidis, P., & Bulbulia, J. (2013). Cognitive resource depletion in religious interactions. *Religion, Brain and Behavior, 3,* 39–55.

Schjoedt, U. and van Elk, M. (in press). The neuroscience of religion. In Barrett, J., editor, Oxford handbook of the cognitive science of religion. Oxford Univerity Press.

Schjoedt, U., Stødkilde-Jørgensen, H., Geertz, A. W., Lund, T. E., & Roepstorff, A. (2011). The power of charisma-perceived charisma inhibits the frontal executive network of believers in intercessory prayer. *Social Cognitive and Affective Neuroscience, 6,* 119–127. https://doi.org/10.1093/scan/nsq023.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47,* 609–612.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage, 44,* 83–98.

Starreveld, P. A., & La Heij, W. (2017). Picture-word interference is a stroop effect: A theoretical analysis and new empirical findings. *Psychonomic Bulletin & Review, 24*(3), 721–733.

Stavrova, O. (2015). Religion, self-rated health, and mortality: Whether religiosity delays death depends on the cultural context. *Social Psychological and Personality Science, 6,* 911–922.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology, 15,* e2000797.

Tang, J., Critchley, H. D., Glaser, D., Dolan, R. J., & Butterworth, B. (2006). Imaging informational conflict: A functional magnetic resonance imaging study of numerical stroop. *Journal of Cognitive Neuroscience, 18,* 2049–2062.

Teper, R., & Inzlicht, M. (2012). Meditation, mindfulness and executive control: The importance of emotional acceptance and brain-based performance monitoring. *Social Cognitive and Affective Neuroscience, 8,* 85–92.

Treiber, J. M., White, N. S., Steed, T. C., Bartsch, H., Holland, D., Farid, N., et al. (2016). Characterization and correction of geometric distortions in 814 diffusion weighted images. *PLoS ONE, 11,* e0152472.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging, 29,* 1310.

van Elk, M., & Aleman, A. (2017). Brain mechanisms in religion and spirituality: An integrative predictive processing framework. *Neuroscience and Biobehavioral Reviews, 73,* 359–378.

van Elk, M., & Snoek, L. (2020). The relationship between individual differences in grey matter volume and religiosity and mystical experiences: A pre-registered voxel-based-

morphometry study. *European Journal of Neuroscience, 51*, 850—865. https://doi.org/10.1111/ejn.14563.

Van Leeuwen, N. (2014). Religious credence is not factual belief. *Cognition, 133*, 698—715.

van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the stroop task: A functional mri study. *Neuroimage, 27*, 497—504.

Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nature Reviews Neuroscience, 6*, 533.

Vuorre, M. (2016). *Meta-analysis is a special case of Bayesian multilevel modeling.* https://vuorre.netlify.com/post/2016/09/29/meta-analysis-is-a-special-case-of-bayesian-multilevel-modeling/.

Wang, S., Peterson, D. J., Gatenby, J. C., Li, W., Grabowski, T. J., & Madhyastha, T. M. (2017). Evaluation of field map and nonlinear registration methods for correction of susceptibility artifacts in diffusion MRI. *Frontiers in Neuroinformatics, 11*, 17.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*, 475—482.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage, 92*, 381—397.

World Values Survey. (2010). *Wave 6 official aggregate v. 20150418.* Asep/JDS Madrid.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods, 8*, 665.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931—959.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging, 20*, 45—57.

Zmigrod, L., Rentfrow, P. J., Zmigrod, S., & Robbins, T. W. (2019). Cognitive flexibility and religious disbelief. *Psychological Research, 83*, 1749—1759. https://doi.org/10.1007/s00426-018-1034-3.

Zuckerman, M., Silberman, J., & Hall, J. A. (2013). The relation between intelligence and religiosity: A meta-analysis and some proposed explanations. *Personality and Social Psychology Review, 17*, 325—354.