Neural Convergence of Graded Belief and Binary Belief Uncertainty

Martin Fungisai Gerchen[1,2,3], Samantha Ullrich[1,2], Mathis Lessau[3,4], & Peter Kirsch[1,2]

[1]Department of Clinical Psychology, Central Institute of Mental Health, University of Heidelberg/Medical Faculty Mannheim, Mannheim, Germany

[2]Department of Psychology, University of Heidelberg, Heidelberg, Germany

[3]Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany

[4]Husserl-Archiv, University of Freiburg, Freiburg, Germany

Corresponding author:
Martin Fungisai Gerchen
RG Biological Psychology
Department of Clinical Psychology
Central Institute of Mental Health (ZI)
University of Heidelberg/Medical Faculty Mannheim
J5
68159 Mannheim
Germany
Email: martin.gerchen@zi-mannheim.de
Tel: +49 621 1703 6163
Fax: +49 621 1703 6505

Abstract

Belief was defined by William James as “the psychological process or function of cognizing reality”, and the recent philosophical literature has emphasized that there are two types of belief: categorical and graded. The relationship between these two belief types is complex, and it is often assumed that the degrees of graded belief reflect confidence. However, this claim has not yet been addressed empirically. In this functional magnetic resonance imaging study we let N=29 young healthy participants rate their graded belief in 16 propositions from the conspiracy theory spectrum and estimated certainty scores from the ratings that we associated with brain activation during the presentation of the propositions. We found associations of uncertainty (i.e. negative associations with certainty scores) with brain activation in the dorsomedial prefrontal cortex (dmPFC), right dorsolateral prefrontal cortex (dlPFC), and posterior parietal cortex. The clusters in the dmPFC and dlPFC replicated associations with explicit ratings of uncertainty in binary belief decisions we identified in a prior study on domain-general belief. These results provide empirical evidence for a link between graded belief and uncertainty in human cognition and emphasize the role of doubt as a relevant process to take into account in concepts of graded belief.

Keywords: belief, uncertainty, dorsomedial prefrontal cortex, conspiracy theories, doxastic attitude, functional magnetic resonance imaging

1 Introduction

Belief is a genuine psychological construct defined by William James as "the psychological process or function of cognizing reality" (James, 1890/1910, Vol. II, p. 283) and its nature has been debated in psychology and philosophy for several centuries (see Gerchen et al., 2024), and more recently also in cognitive neuroscience (Lo Presti et al., 2025). Subjectively held beliefs play an important role in a wide variety of relevant domains from day to day decisions over cognition in mental disorders to political attitudes and endorsement of conspiracy theories (CT) (Bowen et al., 2023; Haghtalab et al., 2021; Kube, 2023; Zhu et al., 2018).

In the recent philosophical literature two types of belief have been contrasted: A categorical binary (true/false) type, which is often extended by a neutral stance as a third option, and a graded type where beliefs can be ordered based on their 'strength', which has also been called degreed doxastic attitude or partial belief (see for example Dietrich, 2022; Foley, 1992; Hawthorne & Bovens, 1999; Jackson, 2020). With respect to their relationship, there have been intensive discussions whether categorical or graded belief are more fundamental, whether each type even exists, and whether and how they could be integrated into a common model (see for example Dietrich, 2022; Dietrich & List, 2018, 2021; Douven & Rott, 2018; Friedman, 2019; Jackson, 2020; Moon, 2017; Schupbach, 2017).

In an influential paper, Foley (1992) made a reference to John Locke's "An Essay Concerning Human Understanding" (Locke, 1689) and proposed the "Lockean thesis" of belief to link the belief types. In this concept, a rational belief is one for which the subjective rational degree of confidence is sufficiently high above some threshold, and it has become common in the philosophical literature to assume that the strength of a graded belief is reflecting confidence (Hawthorne & Bovens, 1999; Jackson, 2020; Schupbach, 2017). However, as Moon (2017) points out, the equivalence between confidence and belief is an assumption that is intuitive but made without further supporting evidence.

It is therefore a relevant question to elucidate the processes involved in belief and graded belief. From an empirical psychological perspective, both types of belief can be directly measured in humans. Subjective belief can be assessed by a binary (true/false) decision, or as a graded belief for example with a truth rating from 0 – 100%, both of which formats are readily answered by participants. To investigating the neural implementation of the involved cognitive processes in humans neuroimaging approaches like functional magnetic resonance

imaging (fMRI) provide important tools that allows to identify associations of brain activity with relevant (sub)processes and potentially differentiate between them.

In a seminal fMRI study on belief, Harris et al. (2008) presented propositions (short statements) from different categories (for example factual, autobiographical, religious) and let participants judge categorically by pressing one of three response buttons whether they considered a shown proposition true (belief condition), false (disbelief condition), or undecidable (uncertainty condition). All statements were designed to be clearly assignable to one of these three categories. Harris et al. (2008) identified associations in the ventromedial prefrontal cortex (vmPFC) for belief, in the inferior frontal gyrus, right middle frontal gyrus, and the bilateral anterior insula for disbelief, and, most relevant in this context, in the anterior cingulate cortex (ACC) and dorsomedial prefrontal cortex (dmPFC) for uncertainty.

In an adapted version of Harris' experiment we presented propositions from the categories facts, politics, religion, conspiracy theories, and superstition, and let participants rate their belief in each proposition (Gerchen et al., 2024). In contrast to Harris et al. (2008), we separated proposition presentation, categorical true/false belief decision, and certainty judgement in this decision on a 0-100% scale to better distinguish the involved processes. In this study, we replicated the vmPFC belief effect of Harris et al. (2008), found an association in the anterior temporal cortex with disbelief, and identified by parametric modulation analysis distinct networks that were linearly associated with certainty and uncertainty. The uncertainty network was centered at the dmPFC, where effects were maximal. Thus, using a substantially different procedure, this dmPFC effect constitutes a conceptual replication of Harris' uncertainty effect. Given the well-established role of the dmPFC in risk processing (Wu et al., 2021), we interpreted it as a marker of the "subjective epistemic risk" of a proposition (Gerchen et al., 2024). In a subsequent study we investigated these effects in self-referential belief, which plays an important role especially in mental disorders like depression, and were able to replicate the vmPFC belief effect as well as the linear dmPFC belief uncertainty effect for this specific belief domain (Bruns et al., 2025).

Interestingly, the dmPFC has also been implicated by Kaplan et al. (2016) in resistance to belief change. In their study each participant was shown 8 political propositions and 8 non-political propositions with which they had indicated strong agreement. Each proposition was followed by 5 challenges providing a counterargument or evidence against the original proposition. The

higher the activation of the dmPFC was during the presentation of challenges contradicting strongly held political and non-political views of their participants, the smaller was the change in the respective belief. While Kaplan et al. interpreted their finding with regard to the role of the dmPFC in cognitive reappraisal and regulating negative affect, this would also be consistent with an underlying uncertainty effect.

To further investigate these processes, in the current study we used an adaptation of the experimental design of Kaplan et al. (2016). In our study we let participants rate their graded belief in 16 CTs on a scale from 0-100% and presented either 5 (alternative facts) or 5 contradicting (official facts) statements for each in a counterbalanced fashion.

While the main aim of our whole study was to identify brain processes associated with processing of information about CTs and their relationship with the Jumping-to-Conclusion (JTC) bias, in this perspective paper we focus solely on a pregregistered secondary analysis to identify brain activation associated with belief uncertainty and use this information to investigate the underlying neural processes of graded belief. This scientific question is independent of the primary goal of the study. The preregistration further included the transformation of the obtained graded belief ratings to (un)certainty scores (see methods section for further details).

In order to compare uncertainty effects estimated from graded belief to uncertainty effects for binary belief decisions and address the question of their similarity, we present our current results in conjunction with those reported in Gerchen et al. (2024).

2 Methods

2.1 Sample

A group of 30 healthy volunteers participated in the study. Exclusion criteria were no MRI eligibility, neurological or acute psychiatric illness, a history of substance dependence other than nicotine, left-handedness and insufficient German language skills. One person was excluded from the analysis because of excessive head motion during fMRI scanning. Thus, the final sample size was N=29 participants (mean age 24.97 years (SD = 4.48, range 19-36 years), 22 female). 19 participants were students, 7 had an academic degree and 3 had a completed apprenticeship. Before participants were included, a screening telephone interview was conducted. Following successful screening, participants received written and oral information about the study and provided written informed consent. At the end of the study, participants were compensated with 20€ or partial course credit. The study was approved by the ethics committee of the Medical Faculty Mannheim, University of Heidelberg (ID 2022-646) and preregistered at aspredicted.org (#120910, https://aspredicted.org/49kz-s6nt.pdf) before the start of data acquisition. All research was conducted according to the principles expressed in the Declaration of Helsinki.

2.2 Paradigm

For the paradigm, we created 16 CT propositions, and for each proposition, 5 confirming (alternative facts) and 5 contradicting (official facts) pieces of information. For example, two of the propositions were "*Die US-Regierung hat den Angriff auf Pearl Harbor bewusst geschehen lassen." (The US government deliberately allowed Pearl Harbor to happen*.) or "*Pharmakonzerne halten wirksame Medikamente zurück um weiterhin Geld zu verdienen" (Pharmaceutical companies hold back effective drugs to keep making money).* In 16 trials, participants were first shown a CT proposition for 6 s and then rated their belief on a visual analog scale from 0 – 100 % with markers '*definitely not true'* (at 0%), '*50/50'* (at 50%), and '*definitely true'* (at 100%) by shifting a slider with a button box. After the rating, they were shown 5 pieces of information (either alternative or official facts for half of the propositions counterbalanced across participants) and a second presentation of the same statement followed again by the belief rating. A fixation cross was shown before each trial.

As preregistered, we calculated certainty scores *c* from graded belief ratings *b* by

$$c = |b - 50| \times 2$$

Irrespective of the direction of the truth rating, certainty c is 0 when belief was rated at 50% (‘*50/50*’) and increases towards c=100 the further the rating approaches the ends of the belief scale. Uncertainty is conceptualized as the negative (inverse) association with the certainty rating, thus, being highest at 50% and decreasing towards the ends of the belief scale.

2.3 MRI Scanning

MRI data was acquired at a 3T Siemens MAGNETOM PRISMAfit scanner (Siemens Healthineers, Erlangen, Germany) at the Central Institute of Mental Health in Mannheim, Germany, in a single session that included an anatomical MPRAGE scan, a 6 minute resting state scan, and the experiment. The MPRAGE was acquired with a time of repetition TR=1.3 s, echo time TE=3.03 ms, flip angle 9°, and isotropic resolution of 1x1x1 mm. Functional images were acquired with an echo-planar imaging (EPI) sequence with TR=1.02 s, TE=30 s, 63° flip angle, simultaneous multi-slice (SMS) factor 2 and GRAPPA factor 2 in 30 slices with a thickness of 3mm and 1 mm gap and 3x3 mm in-plane resolution. The experiment was self-paced, and scanning was stopped manually after completion. The average scanning duration was 19:45 min (SD: 1:03 min; range: 18:18 – 23:02 min). During scanning, respiration and pulse were monitored with built-in equipment and saved.

2.4 MRI Data Analysis

MRI data analysis was conducted with SPM12 (v7738) in MATLAB R2020a. The anatomical image was segmented and normalized to MNI ICBM152 non-linear asymmetric 2009b space. Functional images were slice-timing corrected, realigned, coregistered to the anatomical image and normalized to MNI space by applying the forward deformation field, rescaled to a resolution of 3x3x3 mm, and smoothed with a Gaussian kernel with FWHM=8x8x8 mm.

First level analysis was conducted with a general linear model (GLM) including condition regressors for initial CT statement presentation, its parametric modulation with the estimated certainty scores, repeated CT statements, official facts (blockwise over all 5 information presentations), alternative facts (blockwise over all 5 information presentations), and the rating phase convolved with the canonical hemodynamic response function (cHRF). Furthermore, regressors were included for events of no interest (fixation cross, button presses) convolved with the cHRF, 6 motion regressors, dummy regressors marking volumes

with excessive head motion (estimated with the ART toolbox; parameters framewise displacement FD>0.5 mm, global intensity change z>4), white matter and cerebrospinal fluid signals, and physiological nuisance regressors estimated with the TAPAS PhysiO toolbox (Kasper et al., 2017).

A positive contrast was put on the parametric modulation of the initial statement presentation by subsequent certainty ratings. Second level imaging analysis were conducted on the obtained contrast images with one-sample t-tests with sex and age as covariates. Uncertainty effects were estimated by a negative contrast (-1). The preregistered statistical threshold for whole-brain analysis was cluster-level significance of $p < .05$ with a cluster-defining threshold of $p < .001$ uncorrected.

To compare the identified effects with uncertainty in binary belief we plot the overlap with the respective results of (Gerchen et al., 2024), their Figure 5) at the same statistical threshold (see Figure 1).

3 Results

Participants rated their belief by mean initial statement presentation on *M* = 12.68%, *SD*: 9.76% (total range 0-84%) over all 16 CT statements. A positive second level contrast on the parametric modulation of the initial statement presentation by the estimated belief certainty score showed no significant effect. For the negative contrast (uncertainty), significant clusters were found in occipital cortex, right posterior parietal cortex, right dorsolateral prefrontal cortex, and dorsomedial prefrontal cortex (Figure 1 and Table 1). When plotted on the effects associated with uncertainty ratings for binary belief decisions reported in (Gerchen et al., 2024), overlapping clusters are found in the dmPFC and the right dorsolateral prefrontal cortex (dlPFC, see Figure 1).

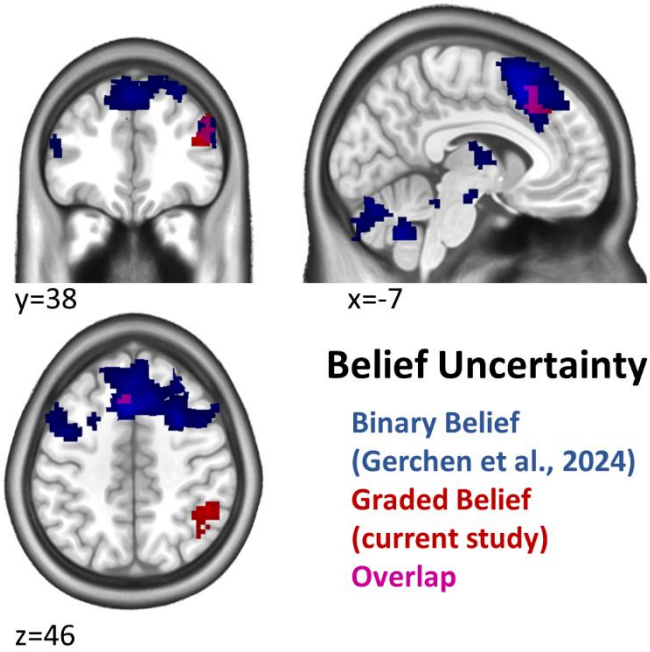

Figure 1. Overlap of graded and binary belief uncertainty. Overlap (magenta) of brain activation during initial proposition presentation associated with uncertainty estimated from graded belief ratings in the current study (red) with brain activation during proposition presentation associated with uncertainty ratings for binary belief decisions reported in (Gerchen et al., 2024) (blue). Significance threshold in both analyses p=0.05 cluster-level corrected with a cluster-defining threshold (CDT) of p=0.001 unc.

Table 1.

cluster-level		**peak-level**			
p_{FWE}	**k_E**	**p_{FWE}**	**t**	**MNI (mm; x, y, z)**	**label**
0.032	63	0.055	6.19	21, -103, 8	occipital cortex
0.001	127	0.294 0.999	5.34 3.86	48, -46, 53 39, -64, 53	posterior parietal cortex
0.045	58	0.488 0.993	5.02 4.00	39, 35, 20 48, 41, 29	dorsolateral prefrontal cortex
0.048	57	0.995 1.000 1.000	3.98 3.78 3.76	-9, 23, 38 -6, 26, 50 -3, 29, 38	dorsomedial prefrontal cortex

Table 1. Effects of parametric modulation of statement presentation with uncertainty.

4 Discussion

We identified a neural signature of belief uncertainty estimated from graded belief that overlaps in the dmPFC and right dlPFC with a signature based on the explicit rating of uncertainty about binary belief decisions. This enables us to link self-reported graded beliefs to uncertainty on the basis of brain processes, which is a relevant insight as an equivalence between belief and confidence had previously been assumed without further evidence in the philosophical literature (Moon, 2020).

While the identification of overlapping brain activations with fMRI cannot be taken as definite proof for the equivalence of underlying cognitive processes (Poldrack, 2006), we think that the robustness of the association provides as much empirical evidence as is possible to derive from noninvasive human brain mapping.

After Gerchen et al. (2024) and Bruns et al. (2025), the reported results constitute our third conceptual replication of the original finding of a belief uncertainty association in the dmPFC by Harris et al. (2008) with another modified experimental procedure. The identification of this association by a categorical uncertainty choice (Harris et al., 2008), an explicit rating of (un)certainty in a binary belief decision (Bruns et al., 2025; Gerchen et al., 2024), and uncertainty estimated from a graded belief rating (this study) emphasizes its robustness and might hint to a fundamental role of the underlying brain process in human belief. While the propositions investigated in the current study were only taken from the domain of conspiracy theories, this suggests that the reported effects are likely expressing domain-general belief processes rather than domain-specific effects (for specific effects in the conspiracy theory domain see Gerchen et al. (2024) and Zhao et al. (2025)).

In addition to the dmPFC we found an overlap in the right dlPFC, which might, together with the cluster in the right posterior parietal cortex we identified only in this study, hint at the involvement of executive functions in belief uncertainty.

The average belief in the presented CT statements was rather low in the investigated sample, which had mainly an academic background. However, the conducted parametric modulation analysis captures activation differences between the presented statements. Thus, it is not required that overall belief in the statements is high when at least some, although limited, variability between items exists. Still, limited variability and the small number of statements

might be reasons why the identified cluster in the current study is smaller than in Gerchen et al. (2024). Despite these limitations we were still able to identify the uncertainty association in the dmPFC, which also in Gerchen et al. (2024) was the strongest effect in the whole data set.

To estimate certainty scores from graded belief ratings, we used a relatively simple equation in which we assumed a linear relationship across positive (belief) and negative (disbelief) directions. This does not necessarily have to be true, and the relationship might be more complex. It would be interesting to investigate the exact form of this relationship in future studies.

Importantly, while the philosophical literature more often emphasizes the role of credence or confidence than of uncertainty in graded belief, we did not find a significant association in this direction, although we identified distinct brain networks that were positively associated with certainty ratings in Gerchen et al. (2024). While this is likely due to low power in the current study, the clear, replicable, and robust association of uncertainty with brain activity speaks for a fundamental role of doubt in human belief processes.

5 Conclusions

It seems natural for humans to rate their beliefs categorically as well as on a graded scale. Our results provide empirical neuroscientific evidence that reported levels of graded belief are, at least to a certain extent, reflecting the degree of uncertainty in a belief object. This corresponds to the notion of Alexander Bain who pointed out that “the real opposite of belief as a state of mind is not disbelief, but doubt, uncertainty” (Bain, 1859, p. 574). Overall, our findings suggest that theoretical conceptualizations of graded belief should focus solely on credence or confidence, but should also consider its inverse direction of doubt or uncertainty, which has a distinct psychological and neural reality.

Data and Code Availability

The datasets generated and analyzed during the current study are not publicly available due to data protection requirements but are available from the corresponding author on reasonable request.

Author Contributions

M.F.G., S.U., and P.K. conceptualized the project. M.F.G. and P.K. supervised the project. M.F.G. and S.U. conducted data acquisition, data curation, and data analysis. M.F.G., S.U., M.L., and P.K. wrote and reviewed the manuscript.

Funding

MFG and ML were supported by the WIN program of the Heidelberg Academy of Sciences and Humanities financed by the Ministry of Science, Research, and the Arts of the State of Baden-Württemberg. The funding sources were not involved in study design, the collection, analysis and interpretation of data, and in the writing of the manuscript.

Declaration of Competing Interests

The authors declare no conflict of interest.

References

Bain, A. (1859). *The emotions and the will*. John W. Parker and Son, West Strand. https://books.google.de/books?id=6DQCAAAAQAAJ

Bowen, T. R., Dmitriev, D., & Galperti, S. (2023). Learning from Shared News: When Abundant Information Leads to Belief Polarization*. *The Quarterly Journal of Economics, 138*(2), 955-1000. https://doi.org/10.1093/qje/qjac045

Bruns, E., Scholz, I., Koppe, G., Kirsch, P., & Gerchen, M. F. (2025, Jan 16). Self-referential belief shares common neural correlates with general belief. *Sci Rep, 15*(1), 2137. https://doi.org/10.1038/s41598-024-84445-6

Dietrich, F. (2022). Categorical Versus Graded Beliefs [Mini Review]. *Frontiers in Psychology, Volume 13 - 2022*. https://doi.org/10.3389/fpsyg.2022.817940

Dietrich, F., & List, C. (2018). FROM DEGREES OF BELIEF TO BINARY BELIEFS. *The Journal of Philosophy, 115*(5), 225-270. https://doi.org/10.5840/jphil2018115516

Dietrich, F., & List, C. (2021). The Relation between Degrees of Belief and Binary Beliefs: A General Impossibility Theorem. In I. Douven (Ed.), *Lotteries, Knowledge, and Rational Belief: Essays on the Lottery Paradox* (pp. 223-254). Cambridge University Press. https://doi.org/10.1017/9781108379755.012

Douven, I., & Rott, H. (2018). From probabilities to categorical beliefs: Going beyond toy models. *Journal of Logic and Computation, 28*(6), 1099-1124. https://doi.org/10.1093/logcom/exy017

Foley, R. (1992). The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly, 29*(2). https://www.jstor.org/stable/20014406

Friedman, J. (2019, 2019/06/01). Inquiry and Belief. *Noûs, 53*(2), 296-315. https://doi.org/https://doi.org/10.1111/nous.12222

Gerchen, M. F., Glock, C., Weiss, F., & Kirsch, P. (2024, Jul). The truth is in there: Belief processes in the human brain. *Psychophysiology, 61*(7), e14561. https://doi.org/10.1111/psyp.14561

Haghtalab, N., Jackson, M. O., & Procaccia, A. D. (2021, May 11). Belief polarization in a complex world: A learning theory perspective. *Proc Natl Acad Sci USA, 118*(19). https://doi.org/10.1073/pnas.2010144118

Harris, S., Sheth, S. A., & Cohen, M. S. (2008, Feb). Functional neuroimaging of belief, disbelief, and uncertainty. *Ann Neurol, 63*(2), 141-147. https://doi.org/10.1002/ana.21301

Hawthorne, J., & Bovens, L. (1999). The preface, the lottery, and the logic of belief. *Mind, 108*(430), 241-264. https://doi.org/10.1093/mind/108.430.241

Jackson, E. G. (2020, 2020/06/01). The relationship between belief and credence. *Philosophy Compass, 15*(6), e12668. https://doi.org/10.1111/phc3.12668

James, W. (1890/1910). *The Principles of Psychology*. Macmillan. https://books.google.de/books?id=4GPSD2Oc5_YC

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016, Dec 23). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Sci Rep, 6*, 39589. https://doi.org/10.1038/srep39589

Kube, T. (2023, Jul). Biased belief updating in depression. *Clin Psychol Rev, 103*, 102298. https://doi.org/10.1016/j.cpr.2023.102298

Lo Presti, S., Bonavita, M., Piga, V., Lozito, S., Doricchi, F., & Lasaponara, S. (2025, Jun). "Don't stop believing" - Decoding belief dynamics in the brain: An ALE meta-analysis of neural correlates in belief formation and updating. *Neurosci Biobehav Rev, 173*, 106153. https://doi.org/10.1016/j.neubiorev.2025.106153

Locke, J. (1689). *An essay concerning human understanding* (Vol. 3). Oxford University Press. https://doi.org/10.2307/2175691

Moon, A. (2017). Beliefs do not come in degrees. *Canadian Journal of Philosophy, 47*(6), 760-778. https://doi.org/10.1080/00455091.2017.1320201

Poldrack, R. A. (2006, Feb). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci, 10*(2), 59-63. https://doi.org/10.1016/j.tics.2005.12.004

Schupbach, J. N. (2017). Troubles for Bayesian Formal Epistemology?: A Response to Horgan. *Res Philosophica*. https://doi.org/10.11612/resphil.1652

Wu, S., Sun, S., Camilleri, J. A., Eickhoff, S. B., & Yu, R. (2021, 2021/08/01/). Better the devil you know than the devil you don't: Neural processing of risk and ambiguity. *NeuroImage, 236*, 118109. https://doi.org/10.1016/j.neuroimage.2021.118109

Zhao, S., Wang, T., & Xiong, B. (2025, May 26). Neural correlates of conspiracy beliefs during information evaluation. *Sci Rep, 15*(1), 18375. https://doi.org/10.1038/s41598-025-03723-z

Zhu, C., Sun, X., & So, S. H. (2018, Mar). Associations between belief inflexibility and dimensions of delusions: A meta-analytic review of two approaches to assessing belief flexibility. *Br J Clin Psychol, 57*(1), 59-81. https://doi.org/10.1111/bjc.12154