Introduction

As the everyday experiences of many humans indicate, we easily fall under the illusion that someone is present in our surroundings: we take cracking branches to be the sound of footsteps, and we see monsters in piles of clothes at night. These experiences are often accompanied by feelings of uncertainty, anxiety, or fear. But do these emotions emerge in response to the illusory presence of strange agents or could it be the other way around? Are we more prone to detect illusory agents around us *when* we are afraid of them?

While this is an interesting question as such, the answer could tell us much about the mechanisms underpinning some religious and spiritual phenomena. Some research has suggested that agency detection – a concept that encompasses different abilities related to detecting agents and their actions in the environment – is linked to supernatural beliefs and extraordinary experiences involving supernatural agents. Most famously, it has been suggested that humans are inclined towards beliefs in supernatural agents due to innate anthropomorphising tendencies (Guthrie, 1993) or because we are equipped with an adaptive, hyperactive cognitive module specialised in agency detection called the Hyperactive Agency Detection Device (HADD; see, e.g., Barrett & Lanman, 2008). Crucially for religion, it was theorised that HADD underlies experiences of a supernatural presence, facilitating the acquirement of supernatural beliefs (see, e.g., Barrett, 2000). However, experimental studies yielded inconclusive evidence of whether humans truly possess HADD. Importantly, Andersen (2019) has recently proposed a new approach to agency detection based on the theoretical framework of predictive processing. In this framework, our domain-general brains do *not* possess any innate specialised modules – including the one for agency detection. Instead, perceptions of (supernatural) agents result from low-reliable sensory data interacting with learnt expectations about the distribution of agents around us.

Still, some recent studies (embedded within the original modular concept of agency detection) have found that people are more inclined toward illusory agent perceptions when they feel threatened (Nieuwboer et al., 2014; van Elk, 2021). While these findings align with the idea of HADD and could explain some encounters with supernatural beings (especially frightening ones, e.g., ghosts), they are quite puzzling from the predictive processing perspective. Why should the feeling of threat increase the number of false agency perceptions if the expected number of agents is stable and we do not possess any specialised agent-related bias causing such an increase? In a predictive processing account, the higher the perceived prior probability of encountering an agent, the higher the chance of experiencing illusory agency. Conditions of threat (i.e., a qualitative change in expectations) should not, by default, increase the subjective prior probability, so the increase in agency detection (i.e., a quantitative change in the number of illusory perceptions) found in some studies calls for further investigation. Therefore, we are interested in the question of whether participants will detect more illusory agents while they feel threatened by them. The answer might move us toward a more thorough understanding of experiences involving supernatural agents believed to be hostile, from ghosts and phantoms to monsters and aliens.

Since directly threatening the participants by hidden dangerous agents in a real-life scenario would be ethically problematic (or the overall believability of our priming would be very low), we decided to use virtual reality as a previously proven way to simulate real-life scenarios without bringing up serious ethical issues. Virtual reality simulating a real environment can successfully induce illusory agency detection, as evidenced by Andersen et al. (2019) and Tratner et al. (2020). Other agency detection studies are following this trend (Maij et al., 2019; van Elk, 2021), with virtual reality being generally considered an ecologically valid tool for the study and induction of agency detection (Andersen et al., 2019; Maij & van Elk, 2019).

Thus, we designed the present study to empirically tackle the problem left open by previous studies: in a sensorially unreliable, agentless environment (a dark, foggy virtual forest), participants instructed to expect a stable number of agents across groups will receive differing descriptions of these agents as either being hostile (experimental group) or neutral (control group). Thus, we test a hypothetical causal effect of the feeling of threat on the number of false positive agency detections. Our goal is not to test the original predictive processing account (which we consider our standpoint), and the null results of our study would not contradict its basic predictions. However, finding an effect of manipulation on the number of false positive detections – i.e., the number of detections will significantly increase in the experimental group – might suggest a more sophisticated picture of the agency detection phenomenon, possibly involving some form of evolutionarily driven bias after all (which might, but not necessarily is realised by HADD). Thus, our study aims to take a first step towards investigating of which model of neurocognitive realisation of agency detection would be promising to pursue in further research. Consequently, the study will also be relevant to assess the strength of a prominent theory that aims to explain the etiology of beliefs about supernatural agents. Overall, the results might help us learn more about the origins of both religious experiences and supernatural beliefs that the mechanisms of agency detection were said to facilitate.

Agency detection and feeling of threat

Before discussing the rationale and design of our study, let us introduce the structure of relationships and problems linking religion, agency detection, and feeling of threat more explicitly. When it comes to the idea of agency detection, its origins are usually linked to the work of Stewart Guthrie (1993), who proposed that in an inherently ambiguous world, it is best for humans to interpret the data around them using a maximally meaningful and complex pattern, namely, the pattern of human behavior. According to Guthrie, humans are naturally inclined to anthropomorphise their surroundings (e.g., objects, phenomena and nonhuman animals), and religious beliefs originate from this inclination. Following the ideas of Guthrie, Barrett & Lanman (2008) proposed that humans are evolutionarily equipped with a specialised HADD module that makes us biased towards over-detecting intentional agents and agency cues in our surroundings. According to the logic behind HADD, it is adaptive for us to falsely assume that a (possibly dangerous) agent is present when the data are ambiguous because assuming otherwise could create the risk of becoming the unseen agent's prey. From this logic, it was derived that while feeling threatened, humans should detect more illusory agents (Maij et al., 2019). However, unlike Guthrie's anthropomorphism, HADD has not been said to be the root of religious belief. Instead, HADD was supposed to be a mechanism that makes humans prone to acquire and transfer supernatural beliefs (that already exist in culture). As Barrett & Lanman proposed, HADD could facilitate experiences that can be accounted for – only or most easily – using supernatural agents' concepts, making individuals more prone to believe in these concepts and engage in their further transfer.

To this day, many studies have been conducted to test whether there exists a relationship between supernatural beliefs and the tendency to over-detect agents (e.g., van Elk, 2013; van Elk, 2015; Tratner et al., 2020) and under what conditions agency detection is more active (e.g., Nieuwboer et al., 2014; Maij et al., 2019). However, these studies have yielded limited support for the HADD model (see Andersen, 2019; Van Leeuwen & van Elk, 2019). This fact, combined with objections based on neuroscientific evidence (Lisdorf, 2007), led to the turn in paradigm towards a novel agency detection model embedded within the predictive processing framework.

The predictive processing framework, which is increasingly popular in cognitive neuroscience (see Friston, 2018), has already been applied to many problems of cognitive science (see Hohwy, 2020) and Cognitive and Evolutionary Science of Religion (Rigoli, 2020; Schjoedt et al., 2013; Taves & Asprem, 2017). Crucially, it was also used by Andersen (2019) in what we will call the predictive processing model of agency detection (PPAD). In the PPAD, agency detection is not seen as a function of a built-in, hardwired module, as the HADD model supposed. Instead, since predictive processing assumes that our brain is a domain-general predictive machine that calculates probabilities of outside phenomena's occurrences, humans should detect agency only if the calculated probability of an agent's presence is sufficiently high. Specifically, the brain estimates the so-called posterior probability of a given hypothesis (such as "an agent is present") based on the perceived prior probability of that hypothesis (i.e., the probability of an agent's presence estimated independently of incoming sensory data and based on previous personal experience or socio-cultural learning) and on the likelihood of that hypothesis (i.e., how well an agent's presence explains the incoming sensory data) (see, e.g., Hohwy, 2013; 2020). Hence, as Andersen (2019) argues, there are no reasons why humans should, *by default,* detect illusory agents[1]. According to the PPAD, we falsely perceive agents only when the individual prior probability estimate of their presence is high and when the environmental data are unreliable. This is because our brain depends primarily on its prior expectations when the likelihood of different predictions is approximately equal (Andersen, 2019). This central prediction of PPAD already found some evidence in a recent experimental study, where participants who were primed to expect more agents in an agentless virtual reality environment reported more illusory agent detections than participants who were primed to expect fewer agents. Increasing the sensorial unreliability also increased illusory agency detection (Andersen et al., 2019).

Crucially, PPAD has been proposed as a robust explanation of extraordinary experiences involving an encounter with a supernatural agent. As Andersen (2019) argues, religions provide us with approximate distributions of supernatural agents in our surroundings (e.g., they inform us where to meet spirits or in what conditions one can hear God), shaping our prior expectations about the probability of agents' nearby presence. Combined with sensorially unreliable or scarce data, often found in religiously meaningful environments (e.g., dim lightning, smoke, and darkness), these expectations can trigger illusory supernatural agency detections. In other words, expecting to encounter a ghost in a dark forest, where discriminatory capacities of the predictive mind are impaired, might result in an illusory perception of such a ghost. Furthermore, personal experience of a ghost sighting could reinforce already possessed expectations and could evoke these expectations in other people via cultural transfer (Andersen,

[1] On a side note, there is one study that seemingly shows that people detect agency in non-agentic objects without being primed to do so. In the classic Heider & Simmer (1944) experiment, participants detected agency in moving triangles without any corresponding priming. However, Radvansky & Zacks (2014) note that Heider & Simmer were quite clear that they have constructed the triangle animation so that it conveys a story and agency cues. Thus, the study does not show that we are prone to detect agency by default, but rather that we easily interpret nonhumanoid objects as agents when they show signs of intentionality.

2019; also suggested by Barrett & Lanman, 2008 in the context of HADD). This idea could explain some of the "third-kind" encounters found in religious contexts not only as prototypical events (such as, e.g., Saul hearing the voice of God, Muhammad meeting the archangel Gabriel, etc.), but as phenomena that happen on a common basis: e.g., according to Bader et al. (2010), as much as fifth of Americans claim to have witnessed a ghost and 10% believe that they have experienced a UFO sighting.

In summary, while agency detection has been linked to supernatural beliefs and experiences, there are at least two competing views on the neurocognitive realisation of its cognitive function: the HADD view and the PPAD. Apart from the differences discussed above, these two models vary regarding their consideration of the evolutionary basis of agency detection. While in the domain-general PPAD, this basis is largely downplayed in favour of socio-cultural learning, the central assumption of HADD is that such specialised module evolved as a safety mechanism governed by the logic that it is better to detect possibly dangerous illusory agents than assume safety and fall their prey (Barrett, 2000; for a critique regarding HADD's possible maladaptiveness see, e.g., Lisdorf, 2007). Based on this idea, Maij et al. (2019) predicted that, assuming the existence of HADD, the feeling of threat should increase illusory agency detection. In PPAD, on the other hand, no such hypothesis is easily derived from the theory, as the feeling of threat does not necessarily alter the perceived probability of encountering an agent. Hence, we should not, by default, expect the feeling of threat to increase the number of illusory agency detections.

According to our best knowledge, results from three sets of studies on this matter seem quite puzzling. In their project, Maij et al. (2019) – more in line with PPAD than HADD – found no effect of the feeling of threat on agency detection across five studies. However, the authors recognise a major methodological limitation of their experimental settings. Namely, the feeling of threat was not induced as directly related to suggested agents; instead, participants were exposed to threatening stimuli before the agency detection task itself. Therefore, threat manipulation and agency detection task were mutually unrelated. On the other hand, other disclosed (Nieuwboer et al., 2014) and non-disclosed (van Elk, 2021) studies conversely reported that feeling of threat increases agency detection. However, these experiments were also conducted with threat priming "detached" from the agency detection task. We believe that these studies left an opening that we want to address with a revised methodology. Thus, in the experimental study, our aim is to address the unresolved question of whether the feeling of threat, directly related to suggested agents, increases agency detection.

The present study

We plan to follow the paradigm started by Andersen et al. (2019) and measure illusory agency detection in an agentless virtual reality environment. While participants in the experimental group will be primed with information about the hostile nature of the agents inhabiting the virtual environment, in the control group, agents will be introduced as neutral. Crucially, we intend to keep the expected number of agents stable across groups, and we assume that our priming and control stimuli will be of approximately the same specificity, inducing priors of approximately the same precision: namely, the likelihood of the same data under these two predictions will remain roughly the same. While we cannot rule out the possibility that the priors in each group will have different precision, we believe that addressing this issue is a matter for further research and does not lie within the scope of this project.

Hence, in our study, we intend to create a straight semantic link between threat priming and the agency detection task, directly addressing the limitation of previous studies. Straight threat priming should allow us to determine whether there exists a bias towards agency detection when one receives information that the agents themselves are dangerous (instead of priming participants with non-specified threatening stimuli). As sensorial unreliability has been found to increase agency detection (Andersen et al., 2019), we intend to keep it a constant condition across all participants to induce false positive detections successfully.

The implications of our study will be relevant to the development of the PPAD model and through it, to our understanding of religious experiences the model is said to underlie. In the context of PPAD, there are no reasons to expect that priming subjects with the threatening nature of agents (a qualitative change in expectations) should lead to an increase in false positive agency detections (a quantitative change in perception). We believe that if such a result is found, a reconsideration of the possible evolutionary aspects of PPAD would be called for. Besides bringing back a modular approach to agency detection, this could be done by assuming that there might be agency detection biases existing in the predictive architecture of the mind (e.g., "evolved priors" or "hyperpriors" that increase the perceived prior probability of encountering an agent given that the agent can be dangerous, see Maij & van Elk, 2019; Asprem, 2019). We are more inclined to develop the possible hyperprior account in case positive results are found, because – looking at a bigger picture – many other studies have shown that the idea of HADD is questionable. However, we will discuss both possible paths and their advantages and disadvantages.

On the other hand, finding no effect of experimental manipulation would mean that we could not find evidence favouring the idea that a hypothetical HADD is triggered by a feeling of threat. In general, our study contributes to resolving some doubts about the mixed results of previous studies on the feeling of threat and agency detection. However, it is not designed to rule out one of the two distinct neurocognitive models of agency detection, but rather to constrain the theoretical space for further studies to come. Hence, we believe that the present study makes for a first step towards a "strong inference" (Platt, 1964) research in the field of agency detection. Instead of focusing on testing one particular prediction of a given theory – which is also more than valuable – our experiment points out which of the competing models is more worthy of a further pursue.

We present and discuss the possible impact of our study in Figure 1 below, which contains a hypothetical development of the PPAD model involving hyperpriors, feasible for interpreting the positive results of our study (for a general discussion of possible implementation of modular insights into the predictive processing framework, see, e.g., Drayson, 2017; Asprem, 2019). Regardless of the results, we believe that our study will add to the debate on agency detection and consequently, – considering that agency detection is understood as a building block underlying religious beliefs and experiences – contribute to our understanding of religious phenomena. The impact of our study on understanding religious and spiritual phenomena (especially religious experiences, such as encounters with supernatural beings) will be adressed more thoroughly in the Discussion.

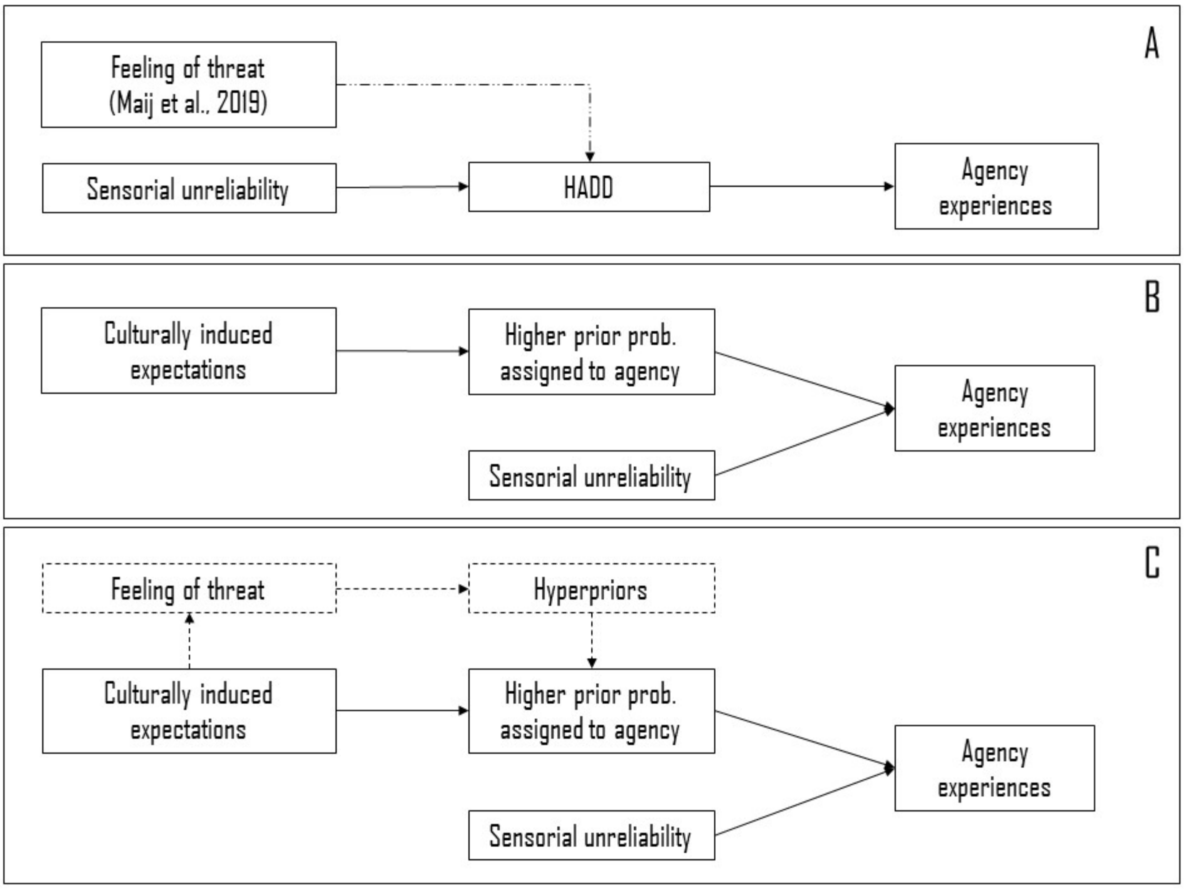

Fig. 1. Possible impact of our study on the development of agency detection models. In figure A, a classic HADD model is depicted: the HADD module, activated by sensorially unreliable stimuli (usually referred to as "ambiguous data" in the original context of the HADD model), triggers agency experiences and can be, but not necessarily always is, also triggered by the feeling of threat (as hypothesised by Maij et al., 2019). Figure A predicts the positive results of our study. Figure B shows a PPAD account, with expectations leading to an increase of the prior probability assigned to the agency, which, combined with unreliable stimuli, leads to agency experiences. Figure B does not predict the positive results of our study. In figure C, we have presented a hypothetical development of the PPAD model, where induced expectations lead to a feeling of threat, under which a hyperprior increases the prior probability assigned to the agency. Figure C is feasible for interpreting positive results of the study instead of returning to the HADD model – a view that we will argue for in the Discussion, should positive results be found.

To our knowledge, we are the first to propose an experiment involving the kind of manipulation discussed above. Since we designed the study to provide novel empirical data and aim to tackle the theoretical shift in understanding neurocognitive realisation of agency detection (where both the null and positive results would be important), we see the preregistration of our study as essential. The format of the registered report ensures that we will publish our data regardless of the nature of the results and increases the probability that the study will motivate further discussion in the field of agency detection.

Method

Study rationale

In the present study, we intend to induce the feeling of threat by using direct semantic priming embedded within the instructions for participants, describing the hostile vs. neutral nature (for the experimental and control group, respectively) of the agents supposedly inhabiting a sensorially unreliable virtual environment. The instructions act as a proxy for culturally induced expectations, analogous to the study by Andersen (2019), but instead of altering the expectations about the probability of encountering an agent (as was done by Andersen, with one instruction suggesting a large number of agents in the virtual reality, and the other – a small number), our instruction for the experimental group is intended to alter the expectations about the nature of agents, while keeping the expected number of agents stable across both groups. After receiving the instructions, the participants will explore a foggy, virtual reality forest (with fog as a proxy of sensorial unreliability) and press a button whenever they feel they have detected an agent. There will be no actual agents in the environment, so all reported detections will be false alarms. The main hypothesis is that participants primed with threatening information about the hostile nature of agents will report more agency detections during the exploration of the environment. A manipulation check with the use of an anxiety questionnaire will be conducted to decide whether our priming was effective.

We will also administer a heart rate measurement, which was previously used as an indicator of fear (e.g., Andersen et al., 2020). Nevertheless, we know that heart rate can also measure other affective states, such as stress related to a gaming-like challenge (e.g., Trotman et al., 2018). Considering this, in our study, the heart rate levels will act as a continuous variable reflecting the general arousal of participants, followed by the check if arousal has a different impact on agency detection in participants from the experimental group compared to controls. Finding such an effect would indicate that our priming has induced specific arousal in the experimental group (most likely, related to feeling threatened by the agents). Hence, our secondary hypothesis is that priming will moderate the relationship between heart rate and agency detection. Specifically, we expect that the effect of heart rate on agency detection will be stronger for participants from the experimental group.

Finally, we will administer a questionnaire measure of anxiety which (apart from being used as a manipulation check) will be tested as a mediator of the effect of priming on agency detection. Both tests will bring us closer to the answer if our priming not only induced anxiety but also that the anxiety made the participants detect more agents in the experimental group. Hence, our third hypothesis is that the effect of manipulation on agency detection will be mediated by anxiety, with higher anxiety in the experimental group leading to increased agency detection.

Regarding our choice of virtual reality, we follow the growing trend of using this method in agency detection studies. In addition to the study by Maij et al. (2019) and van Elk's (2021) non-disclosed empirical investigation, virtual reality has also been successfully used to induce false agency detections by Tratner et al. (2020) and Andersen et al. (2019). Overall, virtual reality is considered an ecologically valid tool for the induction of agency detection (Andersen et al., 2019; Maij & van Elk, 2019).

Hypotheses

In the present study, we intend to experimentally test three hypotheses. Each of these hypotheses is directional (with its alternative written in the brackets). Corroborating these hypotheses is in line with HADD and the developed PPAD model that includes hyperpriors but not with the original PPAD. Finding no evidence for the effect, on the other hand, brings no evidence against PPAD and no evidence supporting HADD or the PPAD model that includes hyperpriors.

The hypotheses are the following:

I. In a sensorially unreliable virtual environment, the feeling of threat induced in the experimental group by semantic priming about the hostile nature of suggested agents will significantly increase false positive agency detections compared to the control group (null hypothesis: the feeling of threat will not increase agency detection).

II. The difference in agency detection between participants with low and high arousal levels will be greater for participants in the experimental group compared to the control group (null hypothesis: the group will not moderate the relationship between heart rate and agency detection).

III. Participants from the experimental group will exhibit higher anxiety levels than the controls, leading to increased agency detection (null hypothesis: anxiety scores will not mediate the effect of priming on agency detection).

The relevant variables will be operationalised as follows:

- The agent-related feeling of threat will be induced with semantic priming embedded within the participant's instruction, acting as a proxy for culturally induced information.
- The effect of manipulation, namely, the level of anxiety felt during the task, will be measured via a modified PANAS-X affect scale (Watson & Clark, 1999).

- Agency detection will be measured through the number of participant's button presses during exploring the (agentless) virtual environment. Participants' task will be to press the button whenever they detect an agent.
- The level of arousal will be measured using a Zephyr Echo BioHarness heart rate measurement tool.

Thus, the operational hypotheses are:

1.
    a. Participants in the experimental group will score higher on the PANAS-X affect scale than participants in the control group (manipulation check).
    b. Participants in the experimental group will press the button more times than participants in the control group.
2. The difference in the number of presses between participants with low and high heart rates during the task will be greater for experimental group participants than for control participants.
3. Participants from the experimental group will press the button more times than participants in the control group, but that effect will disappear or will be significantly lower when controlling for PANAS-X affect scores.

The Masaryk University Research Ethics Committee has approved the design of the study.

Sample

We aim to collect data from 219 healthy volunteer adults (see analysis of power below) who signed informed consent to participate in the study. We use convenience sampling. Recruitment will be done via a pre-existing database of the HUME Lab facility, where the trials will take place, and via social media announcements. We will not include participants with (not corrected) vision or hearing problems, heart diseases, psychological and psychiatric disorders, previous experience of virtual reality side effects and high baseline anxiety levels found in an administered pre-study scale. The participant reward is 150 CZK (approx. $6), and the study will be conducted in the Czech language.

Analysis of power

We performed power analyses for all tests we plan to use to establish our sample size. First, we estimated the effect size based on previous studies to calculate a suitable sample size. Second, we performed classic G*Power analysis. Finally, we conducted an additional

ShinyApp power analysis to determine the minimal difference between means we could find in the study.

Our estimate of the effect size is based on three sources that tested the relationship between the feeling of threat and agency detection:

- [van Elk, 2021]: The author reports the results of five studies testing the influence of threat on agency detection. In three studies, participants attributed more agency to threatening pictures (the effect sizes (η^2) were 0.47, 0.74 and 0.47). In the remaining two, participants reported agency detection often in a threatening (vs non-threatening) virtual environment (although only in one of those, the effect was significant, with $\eta^2 = 0.24$). The latter seems more relevant to us as a point of reference. We calculated the weighted arithmetic mean with a weight of 1 for the effect sizes from the three first studies and a weight of 3 for the effect sizes from the virtual reality study. This way, we have estimated an η^2 of 0.4.
- [Nieuwboer et al., 2014]: The authors reported that participants attributed significantly more agency to threatening pictures than non-threatening ones ($\eta\rho^2 = 0.28$). This effect was found both in pictures of agents ($\eta\rho^2 = 0.21$) and natural phenomena ($\eta\rho^2 = 0.15$). We decided to calculate an arithmetic mean of these three values and estimated an $\eta\rho^2$ of 0.21.
- [Maij et al., 2019]: Maij et al. estimated the effect size for their study using an earlier study by van Elk (2013) with a similar paradigm. The estimate was $\eta\rho^2 = 0.13$. However, in their study, no effect was found, which might be because the effect size in the population is lower than that estimate. To address this issue, using the rule of thumb for interpreting partial eta-squared (small – 0.01; medium – 0.06; large – 0.14) we decided to decrease the assumed effect size by one level: from the large size assumed by Maij et al., to medium size.

Based on these considerations, we estimate the effect size to be medium. Given that estimation, we calculated the sample size required to test our main hypothesis with linear regression using G*Power (Faul et al., 2007). With an effect size of $f^2 = 0.15$ (traditional interpretation of medium size), alpha set to 0.05 and a power of 0.95, the required sample size is 89.

Regarding our moderation and mediation hypotheses (hyp. 2 & 3), we performed power analysis for linear regression with additional predictors (moderator and mediator). We have no estimates of the effect sizes, so we assumed a small to medium effect of $f^2=0.08$. With three predictors in the regression model, the required sample size increases to 219.

Therefore, our baseline number of participants is 219. When it comes to manipulation check, we have no estimates of the effect size, but for an independent samples t-test (one-tailed), with a sample size of 219 and for different effect sizes, the range of test powers is 0.43 for d=0.2 (small effect), 0.98 for d=0.5 (medium effect) and almost 1 for d=0.8 (large effect). Thus, the N=219 is likely to be sufficient, given at least a medium effect. However, since our budget allows us to test as many as 269 participants and some of the participants might not finish the trial (see below), we might run additional trials to collect analysable data from 219 participants. We will not investigate or analyse the data until that number is reached, and we will not conduct additional trials after the count has been reached.

Finally, we ran Shiny App simulations (500 simulations per design; https://shiny.ieis.tue.nl/anova_power) for the main hypothesis with different values of means. Given the means (and SDs) of 2.70 (.26) and 7.40 (.67) found in the study reported by van Elk (2021), the power of the test with 219 participants approximates 1. This is also the case with the same means, but both SDs equal one, and both SDs equal 2. Keeping the SDs at 2 (which gives us a margin for anticipated, higher variance in the data) and with the power of the test not dropping below 0.95, the difference between means can be equal to 1.0 at minimum. Our study is thus sufficiently powered to find an effect with participants in the experimental group detecting approximately one more agent on average compared to the control.

Data exclusion – methodological rationale

During the data collection, we will exclude from the final sample all participants falling into the criteria stated below and run the trials until the maximum number of 219 participants that completed the trial is reached (see also: exclusion based on the statistical rationale in the Data analysis section below):

- (Participant) failed to complete the trial;
- Failed to complete the task in the virtual reality environment (e.g., they took the equipment off during the task, the equipment stopped working, etc.) or were otherwise significantly interrupted during the task;
- Were unable to follow instructions or presented signs of misunderstanding of the basic ideas about the study and the task;
- Presented signs of major side effects of virtual reality or high anxiety;

After data collection, we will also exclude those participants who:

- Correctly guessed that there were no agents implemented in the forest AND guessed it before the task ended (indicated in the final questionnaire);
- Correctly guessed that they were supposed to feel threatened by the agents and detect more agents AND guessed it before the task ended (indicated in the final questionnaire).

Additionally, we will exclude from relevant analyses [indicated in brackets] the data connected to all participants who:

- Had their heart rate measure tool equipped improperly and, as a result, at least 10% of their heart rate scores are missing [all analyses including heart rate]; in all other cases, we will treat the missing data as non-existent, so the average heart rate will be calculated based on the available data.
- Are missing scores in variables: agency detection, relative heart rate increase (difference between baseline HR and HR during the task), PANAS-X responses, PANAS-X full score [all analyses including respective variables].

Materials

Virtual reality equipment and environment

Participants will enter a virtual, dark forest covered in thick fog (proxying conditions of perceptual unreliability) with a faint forest ambience (excluding any agent-related sounds) and the sound of the avatar's footsteps. There is a winding path through the woods that participants ought to follow. Crucially for our study, no agents or agency cues are implemented in the environment. Participants will be using an HTC Vive virtual reality gear. The environment will be created using the Unreal Engine and run on a computer with the following technical specifications: Intel Core i7-5820K 3.30 GHz, 64 GB RAM, 64-bit Windows 10 Pro, and GeForce GTX 980.

Before entering the experimental task, participants will learn how to use virtual reality equipment and the controller in a neutral training area (an empty room with Czech instructions; see Figure 2). Afterwards, participants will explore the forest (see Figure 3) while sitting in a chair, using a controller that allows free motion in all four directions. In both conditions, they will be instructed to press a button on the controller whenever they detect "beings" that are supposed to inhabit the forest.

Fig. 2. Screenshot from the virtual training room – early beta version

Fig. 3. Screenshot from the virtual forest – early beta version

Heart rate measurement

To measure participants' arousal-related heart rate, we will use Zephyr Echo BioHarness equipment. The equipment works on a belt placed on the participant's chest. Heart rate measurement from the 5-minute task will be analysed against a previously measured 5-minute calm baseline. We will subtract the baseline for each participant from the average heart rate during the task. That way, every subject will be assigned a parameter reflecting their relative arousal during the task.

Questionnaires

Following Maij et al. (2019), we have extracted the Fear subscale of the PANAS-X scale (Watson & Clark, 1999), which consists of six items. To these items, we added three more: "anxious" (likewise Maij and collaborators), "threatened" and "tense". On a scale from 1 to 5, participants will indicate to what extent they felt as specified during the exploration of the virtual forest. The reliability of the scale must reach at least $\alpha=0.80$. In case it does not, we will remove, first, the "tense" item, second the "threatened" item and finally, the "anxious" item from the analysis, to bring the data closer to the data from original PANAS-X subscale. If the reliability is still below 0.8, we will perform a confirmatory factor analysis and remove the items that do not fit the model.

Additionally, for exploratory purposes, participants will fill out a short questionnaire at the final stage of the trial, asking (1) what the beings they perceived looked like; (2) if they had any peculiar experiences around any of the button presses; (3) what do they think the real purpose of the study was; (4) if this was their first time using virtual reality; (5) if they regret pressing the button at certain times; (6) about their religious and/or spiritual background. The answer to the third question will be one of the criteria for data exclusion (see the relevant section), and answers regarding religious or spiritual affiliation can be used for additional data analysis. All questionnaires can be found in the Appendix (see “Affect scale” and “Final questionnaire”).

Procedure

Candidates meeting the conditions for participating in the study (see "Sample" section) will be invited to the lab and told that they are participating in a study focused on how perception works in virtual reality. After that, participants will be equipped with a heart measurement tool. The tool will collect baseline calm data for 5 minutes; participants will be instructed to sit calmly and read the informed consent form throughout this time.

Afterwards, participants will be equipped with virtual reality gear and instructed to use it: to look around and move their avatar with the controller. Participants will enter a neutral, training virtual scene to practice further using the gear.

Then, participants will be randomly and double-blindly assigned to the experimental and control group. Both groups receive the final, detailed instructions in the training room just before entering the virtual forest – instructions are delivered automatically from a recording implemented in the software. The basic part of the instruction is the same for both groups: participants are to explore the virtual forest for 5 minutes and focus on their task, namely, detecting the beings that inhabit the forest and clicking the button every time they detect a being. However, the experimental group's instruction contains semantic priming to induce a feeling of threat related to the encounter of (possibly hostile) agents in the virtual environment. In the control group, on the other hand, the beings are introduced as neutral. We attach the English versions of complete instructions for both groups in the Appendix (see “Experimental group instruction” and “Control group instruction”).

After receiving the instructions, all participants will enter the same virtual, foggy forest set in the evening. Following Andersen (2019) and Andersen et al. (2019), we expect fog and darkness to be a proxy of low-reliability sensory data. In the predictive processing framework, the unreliability of sensory data should result in participants' relying on their prior expectations that were directly manipulated by administered semantic priming. During the task, participants will have their heart rate measured.

After finishing the 5-minute exploration of the virtual environment and de-equipping both virtual reality gear and the heart rate measure tool, participants will be asked to fill out the affect scale and the final questionnaire with questions about their agency detection-related experiences, religiosity and the purpose of the study (see section "Materials" above). At the end of the procedure, participants will be debriefed and receive monetary compensation.

Data analysis

The data from the study will be analysed using SPSS, Jamovi, and R software. All data in raw and pre-processed form, as well as all changelogs and analysis scripts, will be uploaded to an open science repository.

Data pre-processing

Before the analysis is conducted, the data will be cleaned of all observations based on methodological exclusion criteria (see above). For every participant, we will calculate their relative increase in heart rate and the sum of their score on the affect scale. We will also test the reliability of the affect scale (Cronbach's alpha model). The number of agency detections will be counted based on the raw times the button was pressed.

In case we find any outliers (+/- 3 SD) in agency detection, relative heart rate increase and PANAS-X full score, we will compare the models with and without including these observations in the analyses, and present both outcomes in the relevant subsections.

Descriptives

We will report the final number of participants after data exclusion, the number of male and female subjects and the average age of the participant (with standard deviation). For variables: age, heart rate, affect, and agency detection, we will present a correlation matrix. The mean and standard deviation will be reported separately for the experimental and control group for heart rate, affect and agency detection.

Hypothesis 1a (manipulation check)

We plan to conduct a simple linear regression (group as the predictor of affect scale scores) to verify whether our manipulation has been successful. All the assumptions for this test will be checked (see subsection below), and in case they are not met, we will conduct an independent samples t-test (after checking the normality assumptions using the Shapiro-Wilk test and homoscedasticity using Levene's test) or Welsch test/Mann Whitney U test (Navarro & Foxcroft, 2019).

Additionally, to control for the effect of virtual reality novelty, we will include the answer to whether this was the first time in virtual reality for a participant as a predictor in the model. If the answer data significantly predicts affect, we will conduct further analyses to determine if virtual reality novelty moderates our effect.

We will report coefficients and significance (for all analysis, $\alpha = 0.05$) and the effect size (adjusted R^2), as well as confidence intervals (CIs; for all analysis, CIs are 95%). If the manipulation check turns out insignificant, we will consider the manipulation unsuccessful and perform only exploratory analyses, reported separately from the main Results section.

Hypothesis 1b (agency detection ~ group)

To test the hypothesis that (1b) participants in the experimental group will press the button more times than participants in the control group, we will conduct a linear regression analysis. For the regression, assumptions of normality of residuals (Q-Q plot), linearity of relationship (eyeball test) and homoscedasticity of residuals (eyeball test) will be checked. We expect that group will predict the heart rate levels to some extent, so we will also check if there is no strong collinearity (VIF < 2.0). If the residuals are not evenly distributed, we will perform a Box-Cox transform following the guidelines from Navarro & Foxcroft (2019). As for the assumption of no collinearity and normality, if a strong correlation is found or the residuals are not normally distributed, we will exchange the model for a t-test, with necessary transformation if the distributions differ from normal and test the respective assumptions.

Additionally, similarly to manipulation check analysis, we will include virtual reality novelty as a predictor in the model and in case of significant effect, we will conduct further moderation analyses.

We will report the results of testing the assumptions, coefficients and significance, effect size (adjusted R^2) and CIs.

*Hypothesis 2 (agency detection ~ HR * M: group)*

To test hypothesis (2) that the difference in the number of presses between participants with low and high heart rate during the task will be greater for participants in the experimental group than for participants in the control, we will perform a PROCESS model of moderation (Model 1; Hayes, 2013), with the heart rate as predictor, group as a moderator and the number of button presses as the dependent variable. We do not expect the relationship between heart rate and agency detection to be non-linear, but should it seem that way, we will switch to non-linear regression. Before conducting the analysis, we will check the respective assumptions (see subsection above), and in case they are not met, we will switch to an ANOVA with interaction. For the moderation hypothesis to be confirmed:

- There must exist a significant interaction effect
- Agency detection must be predicted by HR in the experimental group
- Agency detection must not be predicted by HR in the control group, OR the beta coefficient (slope) is lower in the control group than in the experimental group

We will report the results of testing the assumptions, coefficients and significance, and effect sizes (adjusted R^2) of the whole model, CIs, t-statistics and levels of significance.

Hypothesis 3 (agency detection <- M: anxiety <- group)

To test the hypotheses that (3) participants from the experimental group will press the button more times than participants in the control group, but that effect will disappear or will be significantly lower when controlling for PANAS-X affect scores, we would like to perform a PROCESS model of mediation (Model 4; Hayes, 2013) using bootstrapping (=5000), with the group as a predictor, anxiety as a mediator and the number of button presses as the dependent variable.

For the regression, we will conduct the assumptions test analogous to those as indicated in the subsection above and report their results. For the mediation hypothesis to be confirmed:

- The group must predict agency detection AND the group must predict anxiety AND anxiety must predict agency detection
- The total effect of group and anxiety on agency detection must be significantly lower after subtracting the indirect effect of group (through anxiety) on agency detection.

We will report (all CIs are 95%): the total effect (coefficient, p-value and CIs); the direct effect of group on agency detection (coefficient, p-value and CIs), the effect of anxiety on agency detection (coefficient, p-value and CIs), the effect of group on anxiety (coefficient, p-value and CIs) and mediation (coefficient and bootstrap CIs).

Equivalence testing

To ensure that we can conclude from null results, we plan to conduct equivalence testing if the test of our main hypothesis (1b) turns out to be statistically insignificant. We want to use the TOST (two one-sided tests) procedure with the smallest effect size of interest equal to Cohen's d=0,15. We estimated the SESOI using the "small telescopes" method (Simonsohn, 2015; Lakens et al., 2018): using findings from van Elk (2021) and Nieuwboer et al. (2014), we calculated the effect sizes for a 0.33 test power and the respective sample sizes. The arithmetic mean of these effect sizes, rounded down, is d=0.15, and we consider it the borderline of a meaningful effect in our research. Our alpha level for both one-sided tests is 0.05. We will report the value of t, 90% CIs and the test's significance with the higher p-value.

Study timeline

Below we present a brief timeline of our project.

- March 2022 – June 2022: Study preparations.
- July 2022: Pilot trials.
- August 2022 – March 2023: Review period.
- April 2023 – May 2023: Data collection.
- June 2023: Data analysis.
- July 2023: Writing stage 2 manuscript.
- After July 2023: Presenting findings at a conference.
- By December 2023: Closing the project.

Pilot data

We conducted a pilot study to test our procedure, obtain feedback from participants, and possibly improve our design (N=11; 4 female; 2 non-binary; age: *M*=29.72, *SD*=6.60). We report in the following the general results of these trials; all improvements to our study based on the pilot trials have already been included in the relevant sections above.

Most participants (7/11) detected at least one being in the forest (*M*=2, *SD*=2,60). However, despite some of them clicking the button, in the final questionnaire, 7 participants (partially overlapping with 7 participants mentioned above) reported that they did not detect any agents.

We did not find any significant difference in affect score between groups (two-tailed test; $t=0.43$, $p=0.678$), and the average score in the control condition was higher than in the experimental group, which motivated us to improve our instructions, making them more suggestive.

We also performed a linear regression to test whether the group and the relative increase in heart rate predict agency detection. We calculated relative arousal for every participant, subtracting the average baseline heart rate from the average heart rate during the task. Neither condition ($\beta=-0.52$, $p=0.698$) nor heart rate ($\beta=-0.03$, $p=0.701$) significantly predicted agency detection. Like in the manipulation check, we improved our instructions for the task since the average number of button presses was lower in the experimental group than in the control group.

Appendix

Experimental group instruction

Note: Differences between exp. and control instructions have been indicated in italics.

"You will now *trespass into* a forest inhabited by certain beings. *The forest is their territory, and you are not welcome there*. You can explore the area for 5 minutes – after that, the environment will shut down automatically. While you are free to wander wherever you want, try to more or less follow the path.

During the exploring, you will have one very important task. The task is to try to detect the beings that inhabit the forest. There is about a 95% chance that you will succed. *Be cautious, though. The beings do not like those who enter the forest and* observe and follow every *intruder.* Stay aware of your surroundings and *be very careful around these beings.* As soon as you feel something that might be one of those beings, please press the button as fast as possible. You do not have to be 100% sure, but it is important that you click as fast as possible.

Remember to constantly search for the beings *and be careful.* The beings will be watching you from hiding and *would not like to be detected by you.* They are elusive and evasive – if you do not see them, it does not mean that they are not there! *They would not let an intruder cross the forest unnoticed.* Remember to press the button when you get the slightest feeling that you have detected one of them.

Move forward to enter the forest. Good luck!"

Control group instruction

Note: Differences between exp. and control instructions have been indicated in italics.

"You will now *enter* a forest inhabited by certain beings. *The forest is theirs, but they do not mind your presence there.* You can explore this forest for 5 minutes – after that, the environment will shut down automatically. While you are free to wander wherever you want, try to more or less follow the path.

During the exploring, you will have one very important task. The task is to try to detect the beings that inhabit the forest. There is about a 95% chance that you will succeed. *Do not worry about the beings, though. They are not unfriendly and are not bothered by your presence.* Stay aware of your surroundings and *search closely for these beings.* As soon as you feel something that might be one of those beings, please press the button as fast as possible. You do not have to be 100% sure, but it is important that you click as fast as possible.

Remember to constantly search for the beings *and stay focused.* The beings will be watching you from hiding, and *they will try to remain hidden.* They are elusive and evasive – if you do

not see them, it does not mean that they are not there. *They keep close to any visitor who explores the forest.* Remember to press the button when you get the slightest feeling that you have detected one of them.
Move forward to enter the forest. Good luck!"

Affect scale

"Try to remember what it was like to explore the virtual environment. How did you feel during the walk through the woods? What are the emotions that you felt?
The questionnaire below consists of words that describe feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you have felt this way **during the exploration of virtual environment**. Use the following scale:
1 – very slightly or not at all; 2 – a little; 3 – moderately; 4 – quite a bit; 5 – extremely
AFRAID / SCARED / FRIGHTENED / NERVOUS / JITTERY / SHAKY / ANXIOUS / THREATENED / TENSE"

Final questionnaire

"Before finishing the study, we would like to ask you to answer a couple of questions regarding your experience. Please find them below.
1. Please mark your gender and enter your age: male / female / non-binary / other / I prefer not to tell / AGE
2. Was that your first time using virtual reality? Yes / No
3. Please recall your experience when encountering the beings in the forest. What do you think the beings were? What do you think they look like? Write down your answer below:
4. Do you now regret pressing the button at certain times? How many times do you now believe you actually encountered a being? Write down your answer below:
5. Do you think you can recall any outstanding experiences you had when pressing the button? Did any of the moments when you pressed the button feel especially scary, interesting or in any other way peculiar? Below you can write down, next to the given number of button presses, what kind of experience you had – if you pressed the button more, just add the numbers to the list. If that will help you, you can now ask the Research Assistant to show you the exact timing of each button press during your exploration of the forest – it might refresh your memory.
6. What do you think the real purpose of the study was? Can you recall at what approximate moment you started to suspect the real purpose? If you did not suspect it at all – that is fine.

After completing this questionnaire, you will learn the purpose and why it could not have been explicitly revealed.

7. Are you affiliated with any religious organisation? (e.g., institution, church, brotherhood, movement, etc.) Yes / No / Hard to tell

8. Are you following any form of spirituality? Yes / No / Hard to tell

9. Do you consider yourself to be a religious person? Yes / No / Hard to tell

10. Do you consider yourself to be a spiritual person? Yes / No / Hard to tell

11. Do you engage in any form of religious or spiritual activity? (e.g., ritual, prayer, confession, yoga, meditation, alternative healing, etc.) Yes / No / Hard to tell

12. On a scale of 1 to 7, to what extent do you consider yourself religious? (1=not at all; 7=extremely)

13. On a scale of 1 to 7, to what extent do you consider yourself spiritual? (1=not at all; 7=extremely)

14. On a scale of 1 to 7, to what extent religion feels important to you? (1=not at all; 7=extremely)

15. On a scale of 1 to 7, to what extent spirituality feels important to you? (1=not at all; 7=extremely)

16. Do you believe in any sorts of a supernatural being? (e.g., God, god(s), ghosts, spirits, aliens, demons, angels, etc.) Yes / No / Hard to tell

17. Do you believe in any sorts of supernatural beings that are believed to be merciful or/and friendly? (e.g., merciful god(s), friendly spirits, angels, etc.) Yes / No / Hard to tell

18. Do you believe in any sorts of supernatural beings that are believed to be malicious or/and hostile? (e.g., wrathful god(s), evil ghosts, demonic creatures, etc.) Yes / No / Hard to tell

References

Andersen, M. (2019). Predictive coding in agency detection. *Religion, Brain & Behavior, 9*(1), 65-84.

Andersen, M., Pfeiffer, T., Müller, S., & Schjoedt, U. (2019). Agency detection in predictive minds: a virtual reality study. *Religion, Brain & Behavior*, *9*(1), 52–64.

Andersen, M. M., Schjoedt, U., Price, H., Rosas, F. E., Scrivner, C., & Clasen, M. (2020). Playing With Fear: A Field Study in Recreational Horror. *Psychological Science*, *31*(12), 1497–1510.

Asprem, E. (2019). Predictive processing and the problem of (massive) modularity. Open peer commentary on the target article: Marc Andersen, "Predictive Coding in Agency Detection". *Religion, Brain & Behavior*, *9*(1), 84–86.

Bader, C., Mencken, F., Baker, J. (2010). *Paranormal America: Ghost Encounters, UFO Sightings, Bigfoot Hunts, and Other Curiosities in Religion and Culture*. New York University Press.

Barrett, J. L. (2000). Exploring the Natural Foundations of Religion. *Trends in Cognitive Science, 4*(1), 29–34.

Barrett, J. L. (2010). The relative unnaturalness of atheism: On why Geertz and Markússon are both right and wrong. *Religion*, *40*(3), 169–172.

Barrett, J. L., & Lanman, J. A. (2008). The science of religious beliefs. *Religion, 38*(2), 109-124.

Drayson, Z. (2017). Modularity and the Predictive Mind. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing: 12*. MIND Group.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods 39*, 175-191.

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience, 21*(8), 1019–1021.

Guthrie, S. E. (1993). *Faces in the Clouds. A New Theory of Religion*. Oxford University Press.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis*. Guilford Publication, Inc.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57,* 243–259.

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

Hohwy, J. (2020). New directions in predictive processing. *Mind and Language*, *35*(2), 209–223.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 259–269.

Lisdorf, A. (2007). "What's HIDD'n in the HADD?". *Journal of Cognition and Culture, 7*(3-4), 341-353.

Maij, D. L. R., & van Elk, M. (2019). Evolved priors for agent detection. *Religion, Brain & Behavior*, *9*(1), 92–94.

Maij, D. L. R., van Schie, H. T., & van Elk, M. (2019). The boundary conditions of the hypersensitive agency detection device: an empirical investigation of agency detection in threatening situations. *Religion, Brain & Behavior*, *9*(1), 23–51.

Navarro, D. J. & Foxcroft, D. R. (2019). learning statistics with jamovi: a tutorial for psychology students and other beginners. (Version 0.70). DOI: 10.24384/hgc3-7p15 [Available from url: http://learnstatswithjamovi.com]

Nieuwboer, W., van Schie, H. T., Wigboldus, D. (2014). Priming with Religion and Supernatural Agency Enhances the Attribution of Intentionality to Natural Phenomena. *Journal for the Cognitive Science of Religion, 2*(2), 97-120.

Radvansky, G., & Zacks, J. M. (2014). *Event cognition*. Oxford: Oxford University Press.

Platt, J. R. (1964). Strong inference. *Science, 146*(3642), 347–353.

Rigoli, F. (2020). A computational perspective on faith: religious reasoning and Bayesian decision. *Religion, Brain & Behavior, 2*, 147-164.

Schjoedt, U., Sørensen, J., Nielbo, L., N., Xygalatas, D., Mitkidis, P., Bulbulia, J. (2013). Cognitive resource depletion in religious interactions. *Religion, Brain & Behavior, 3*(1), 39-55.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*, 559–569.

Taves, A. & Asprem, E. (2017). Experience as event: event cognition and the study of (religious) experiences. *Religion, Brain & Behavior, 7*(1), 43-62.

Tratner, A. E., Shackelford, T. K., Zeigler-Hill, V., Vonk, J., & McDonald, M. M. (2020). Fear the unseen: supernatural belief and agency detection in virtual reality. *Religion, Brain & Behavior*, *10*(2), 118–131.

Trotman, G. P., Williams, S. E., Quinton, M. L., & Veldhuijzen van Zanten, J. J. C. S. (2018). Challenge and threat states: examining cardiovascular, cognitive and affective responses to two distinct laboratory stress tasks. *International Journal of Psychophysiology, 126,* 42–51.

van Elk, M. (2013). Paranormal believers are more prone to illusory agency detection than skeptics. *Consciousness and Cognition: An International Journal, 22*(3), 1041–1046.

van Elk, M. (2015). Perceptual Biases in Relation to Paranormal and Conspiracy Beliefs. *PLoS ONE, 10*(6), Article e0130422.

van Elk, M. (2021). What's hidden in my filedrawer and what's in yours? Disclosing non-published findings in the cognitive science of religion. *Religion, Brain & Behavior*, *11*(1), 5–16.

Van Leeuwen, N., & van Elk, M. (2019). Seeking the supernatural: the Interactive Religious Experience Model. *Religion, Brain & Behavior, 9*(3), 221-251.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form.* Cedar Rapids.