Exploring the Association between Paranormal Beliefs and Cognitive Deficits

Charlotte E. Dean

August 2022

Word Count: 51,318

Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the degree of Doctor of Philosophy

School of Life and Medical Sciences

Department of Psychology, Sport and Geography

University of Hertfordshire

ACKNOWLEDGEMENTS

The production of this thesis would not have been possible without the incredible support I have received over the last three years. I would first like to say an immense thank you to my PhD supervisors – Professor Keith Laws, Professor Tim Gale, Dr Karen Irvine, and Dr Shazia Akhtar – for their unwavering support and camaraderie. Their trust and encouragement for me to work independently on this project has challenged me to grow as a researcher and has allowed me to produce work that I am exceptionally proud of. Their expertise and guidance have been instrumental in shaping my academic development, and their support during the many highs and lows of this work has been invaluable. It has been a true pleasure to work with each of them over the course of this project.

Thanks should also go to all of the academics and technical staff in Psychology for fostering such a welcoming community that they have truly made me feel part of during my time at the university. I would like to extend particular thanks to Dr Nicholas Shipp for his friendship and the opportunities he has afforded me to develop my professional career. I am also grateful for every participant who took the time to complete my research studies; this work truly could not have been completed without them.

I would be remiss in not mentioning my family and friends for their belief and encouragement which has lifted my spirits through this process. I would also like to thank the lovely members of Weston RFC for ensuring I remembered to have fun during the long winter months, and for showing a genuine interest in my work.

Finally, I would like to dedicate this work to my grandad, Peter Tilbury, who never saw this adventure but was a constant inspiration throughout my doctoral degree.

"Happiness is not to be found in knowledge, but in the acquisition of knowledge."

-Edgar Allen Poe

PUBLICATIONS

Dean, C.E., Akhtar, S., Gale, T.M., Irvine, K., Grohmann, D., & Laws, K.R. (2022). Paranormal beliefs and cognitive function: A systematic review and assessment of study quality across four decades of research. *PLoS ONE, 17*(5), e0267360.

Dean, C.E., Akhtar, S., Gale, T.M., Irvine, K., Wiseman, R., & Laws, K.R. (2021). Development of the Paranormal and Supernatural Beliefs Scale using classical and modern test theory. *BMC Psychology, 9*(1), 1-20.

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

ABSTRACT

This thesis investigates whether paranormal beliefs are associated with deficits of cognition (in accord with the so-called "cognitive deficits hypothesis"). A systematic review of four decades of research (71 studies, n = 20,993) on paranormal beliefs and cognitive functioning is presented, considering the quality of existing research and identifying areas for future work. It is concluded that study quality in this research area is generally good, although areas of methodological weakness exist including: the lack of preregistration, discussion of limitations, a-priori justification of sample size, and a reliance on undergraduate samples. Heterogeneity of study findings exists, with the most consistent findings emerging for positive associations between paranormal beliefs and both intuitive thinking and confirmatory bias, and a negative association with conditional reasoning ability and the perception of randomness. While most studies suggest a negative relationship between paranormal beliefs and cognitive functioning, the evidence is not convincing of an overall cognitive deficit.

The first empirical chapter (Chapter 3) describes the psychometric assessment of the most widely used paranormal beliefs measure, and the subsequent development of a more up-to-date and reliable measure of paranormal beliefs in the general population (owing to issues surrounding the item content and statistical techniques used to develop existing measures). Two methods of scale development (the classical test theory method of factor analysis, and the modern test theory Rasch analysis) were compared, with the Rasch method providing the most prudent measure of paranormal beliefs in the general population. In particular, the Rasch method allowed for assessment of item difficulty, functionality of the rating scale, and differential item functioning. The final scale measures paranormal beliefs along a single dimension using 13 items and a 4-point Likert scale. Chapter 4 seeks to add to the small volume of research concerning paranormal beliefs and executive functions

identified in the systematic literature review by examining whether paranormal beliefs are associated with executive difficulties. Specifically, the chapter focuses on cognitive flexibility, presenting a series of replication studies to determine the strength and direction of its association with paranormal beliefs. A negative relationship between paranormal beliefs and cognitive functioning was identified in some studies, however heterogeneity was high between studies and an overall weak effect suggestive of external mediating factors. The final empirical chapter (Chapter 5) explores the subjective importance of personal paranormal experiences for the development and maintenance of paranormal (dis)belief. Compared to sceptics, believers provided more detailed memory descriptions, which were classified into five distinct components representing their paranormal experiences: explaining experiences, intuitions, perceptual experiences, spiritual experiences, and explicit memories of experiences. The memory descriptions of sceptics were shorter and more homogeneous than those of believers and were classified into just two classes: justification of disbelief, and experiences with others. Believers overall emphasised individual experiences, while sceptics focused on shared experiences. Personal paranormal experiences were found to be important for both paranormal belief and disbelief. It is suggested that future research place greater focus on the personal experiences of sceptics, which have received little attention to date, using both qualitative and quantitative methods to better understand paranormal disbelief.

While the work presented here identified some cognitive *differences* between paranormal believers and sceptics, little evidence points to an underlying cognitive deficit associated with paranormal beliefs. This suggests that the cognitive deficits hypothesis present within the literature is not an accurate reflection of the relationship between paranormal beliefs and cognition, prompting the need for a new theory and further investigation. Specifically, the work presented here suggests a new fluid-executive theory to test for a potential difference in fluid intelligence and higher order executive functioning that may influence believers' and sceptics' cognitive performance (particularly on tasks requiring novel or abstract problem-solving). The present work also advocates further research focused on quantitative and qualitative relationships between paranormal *dis*belief (scepticism) and cognitive functioning; a perspective that has not received adequate investigation in the literature to date.

CHAPTER 1. GENERAL INTRODUCTION

A brief history of belief in the paranormal

> "There is no form of belief so deeply rooted in man's nature, so widely spread over his entire history…as a conviction of the existence of an unknown and invisible world…" (Agnew & Bidwell, 1863, pp. 41).

The late 19th century is often associated with substantial technological and scientific advances. With these advances, however, came a rise in interest and tolerance towards "extraordinary phenomena", including mesmerism, crystal-gazing, clairvoyance, and parlour magic. Also coinciding with this was the so-called "spiritualist movement", which saw public and private séances performed by mediums that captivated sitters with theatrical manifestations of spirit phenomena. These practices would now be described as "paranormal", a term used to describe any phenomena which contradict the basic limiting principles of current scientific understanding (Broad, 1949).

Modern surveys consistently indicate that belief in these phenomena is still prevalent within the general population. For example, a representative survey of British adults conducted by the market-research company BMG Research (2017) found that a third of their sample believed in paranormal phenomena, and a further 21% were 'unsure'. Of those who either believed in the paranormal or were unsure, 40% indicated they had seen or felt the presence of a supernatural entity. Similarly, Pechey and Halligan (2011) found 30% of participants held at least one strong paranormal belief, and 79% held at least one paranormal belief at any strength (weak, moderate, or strong belief). Comparable levels of belief have been documented across various cultures over recent decades (Pérez Navarro & Martínez Guerra, 2020; Eder et al., 2011; Göritz & Schumacher, 2000; Clarke, 1991). With high levels of interest in these phenomena still consistently reported, it is pertinent to consider why belief in the paranormal persists and what cultivates such belief.

Foundations of paranormal beliefs

Childhood and personal experience

Several distinct avenues of research have focused on explaining both the development and the maintenance of paranormal beliefs. One suggestion is that paranormal beliefs stem from factors of one's childhood, such as high levels of fantasy (Lawrence et al., 1995). Particular emphasis, however, has been placed on negative childhood experiences that are associated with a lack of control. For example, several studies have explored a link between paranormal beliefs and traumatic childhood experiences. Perkins and Allen (2006) found significantly higher paranormal beliefs among those with childhood experiences of physical abuse compared to those without abuse experiences, as well as significantly higher belief in psi, precognition, and witchcraft. This supports findings from Rogers et al. (2007) and Rabeyron and Watt (2010) who found significant positive correlations between paranormal beliefs and self-reported childhood trauma (relating to negative home atmosphere/neglect, sexual abuse, physical punishment), as measured by the Child Abuse and Trauma Scale (CATS; Sander & Becker-Lausen, 1995). While more recent work by Berkowski and MacDonald (2014) found no significant relationship between paranormal beliefs and childhood experiences of physical abuse, significant positive correlations were found for childhood experiences of emotional abuse (with heightened global paranormal beliefs, as well as belief in spiritualism and superstition) and childhood experiences of sexual abuse (heightened global paranormal beliefs and beliefs in precognition, psi, superstition, and witchcraft).

It has been suggested that an assurance of order and meaning in the world is essential for both emotional and psychological well-being, with traumatic events threatening this as they "imply that the world sometimes is uncertain, chaotic, and beyond the individual's understanding and mastery" (see Irwin, 2009, pp. 101-102). In this sense, paranormal beliefs provide the individual with a sense of control by structuring events in such a way that they become comprehensible (Irwin, 2009). This has been supported by empirical work showing significant negative relationships between individuals' perceived childhood control and paranormal beliefs (Watt et al., 2007). Some evidence also supports a relationship between paranormal beliefs and desire for control in adulthood (see Greenaway et al., 2013; Irwin, 2000), suggesting that paranormal beliefs stem from a general lack of control rather than specific childhood experiences that are tied to a lack of control, although this link is not consistently reported (see Blanco et al., 2015; Lobato et al., 2014). Irwin (2009) hypothesised that other, less extreme, childhood experiences can also contribute to heightened paranormal beliefs if they reduce an individual's sense of control, such as being a younger sibling or having authoritative parents. While less evidence has been found for a relationship between birth order and paranormal beliefs, some evidence has supported a link between authoritative parenting (or excessive parental control) and increased paranormal beliefs (see Karayagiz & Aktan, 2020; Rogers & Lowrie, 2018; Watt et al., 2007). Childhood experience of parenting has also been shown to increase paranormal beliefs when parents were spiritual (van Elk, 2017), encouraged imaginative thinking (Makasovski & Irwin, 1999), or demonstrated a positive attitude towards paranormal phenomena (Lindeman & Aarnio, 2006). These findings not only highlight the importance of specific childhood experiences for the development of paranormal beliefs, but also suggest an influence of social environment.

Social influences

In his examination of the ways in which beliefs are formed, Peirce (1884-1886/1931) argued that "unless we make ourselves hermits, we shall necessarily influence each other's opinions". In this vein, several studies demonstrate the impact of social influence on paranormal beliefs. For example, Ridolfo et al. (2010) focused on beliefs in extrasensory perception, and found that individuals are more accepting of this type of paranormal phenomena when beliefs in extrasensory phenomena appear to be popular amongst others. An earlier study, however, employing a similar research paradigm found no relationship between the acceptance of extrasensory perception and the popularity of these beliefs amongst others (Vallee, 2008). Similar inconsistency is seen when examining the specific influence of peers on paranormal beliefs. Some studies have found a positive relationship between paranormal beliefs and peer attitudes towards the paranormal (see Aarnio & Lindeman, 2007; Lindeman & Aarnio, 2006), while others show no significant influence of the level of belief held by participants' friends (but evidence of paranormal belief transmission through the passive social influence of a confederate; Markovsky & Thye, 2001). Some have suggested that factors of one's specific sociocultural environment are more important for fostering paranormal beliefs. For example, Ilori et al. (2014) compared two distinct ethnic groups (Igbo and Yoruba) in the capital city of Ekiti State, Nigeria, and found significant differences in the level of paranormal beliefs between the two groups. When comparing cultural influences across different countries, American samples have been shown to have significantly higher levels of global paranormal beliefs than British samples, as well as increased beliefs in superstition, extraordinary life forms, spiritualism, and precognition (Davies, 1988). Belief in telepathy has been shown to be higher in samples from Iceland and Great Britain than those from Sweden (Haraldsson, 1985), and global paranormal beliefs (as well as belief in spiritual phenomena) have been shown to be higher in samples from Singapore than those from Canada (Otis & Kuo, 1984). The self-reported benefits of paranormal beliefs have also been shown to differ cross-culturally. For example, Burger and Lynn (2005) found that American baseball players had more superstitious beliefs than Japanese baseball players, and reported that their superstitions aided their individual performance, while Japanese players reported that their superstitions benefited the overall performance of the team. This difference in beliefs could be attributed to the cultural emphasis on individualism versus collectivism, respectively.

The historical shift towards individualism seen in Western cultures has been thought to contribute to the persistence of paranormal beliefs, owing to the greater importance given to freedom of choice and personal experiences (see Partridge, 2013). This has also seen a rise in the consumption of paranormal phenomena as services to the self, such as alternative therapies (e.g., reiki), and various forms of live and recorded entertainment. In this sense, paranormal phenomena are often looked on "in much the same way as other purchasable goods" (Partridge, 2013; Roof, 1993, pp. 195). With respect to entertainment, media representations of paranormal phenomena have shown associations with increased paranormal beliefs. For example, resent research by Mowen et al. (2022) found that paranormal media consumption (e.g., television shows, films, comic books, podcasts, etc.) was positively associated with paranormal beliefs. This supports earlier work by Brewer (2012), who found that viewing paranormal reality television shows (e.g., those focused on paranormal investigation), but not paranormal drama shows (e.g., those focused on fantasy-based plotlines), predicted paranormal beliefs. Conversely, media consumption has also been shown to reduce paranormal beliefs. For example, Sparks et al. (1998) took a specific focus on UFO beliefs and found that manipulation of a media story about UFOs affected beliefs, such that stories with discrediting information from scientific authorities reduced UFO beliefs. However, as research in this area often relies on the use of correlational methods of analysis, the causal direction of the relationship between paranormal media consumption and paranormal beliefs is difficult to determine. It is possible that the relationship is a function of paranormal believers being more likely to consume paranormal media than sceptics, rather than paranormal media directly influencing an individual's level of belief. As Zimmer (1985) suggests, individuals may believe in the paranormal because they want to believe and are therefore more likely to involve themselves in paranormal subculture to engage with these beliefs. Sparks and Miller (2001) provide some support for this idea, finding that exposure to paranormal television content only influenced paranormal beliefs if individuals had prior paranormal experiences. These findings, however, directly oppose those reported in earlier work which suggested the relationship between paranormal media consumption and paranormal beliefs is contingent on an individual having *no* prior paranormal experiences (Sparks et al., 1997). While these studies present conflicting findings, they highlight the role that individual differences may play in the relationship between external social influences and paranormal beliefs.

Personality

Many researchers argue that individual differences in personality traits are important to consider in theories of paranormal belief development. Several studies have suggested that an intuitive-feeling personality type, as indicated by the Myers-Briggs Type Indicator (MBTI; Myers, 1962), is positively associated with paranormal beliefs (see Kennedy, 2005; Gow et al., 2003; Brugger & Baumann, 1994; Lester et al., 1987). The intuitive personality dimension is characterised by a preference for novelty, the unusual, grasping the “bigger picture”, and reliance on “sixth-sense” or personal hunches (Fretwell et al., 2013). The feeling dimension emphasises personal and social values in decision-making, empathy, and a preference for subjective information (Fretwell et al., 2013). In this respect, Keirsey (1998, pp. 145) suggests that those with an intuitive-feeling personality type show an interest in occultism and parapsychology in attempt to transcend the material world and gain greater insight into the metaphysical world (cf. Kennedy, 2005). This might also explain the relationship between paranormal beliefs and openness to experience that has been reported in several empirical studies. Work by Kumar et al. (2020) found a positive correlation between openness to experience and belief in extraordinary lifeforms, while Betsch et al. (2020) found openness to experience is predictive of global paranormal beliefs. Williams and Roberts (2016), however, report *negative* correlations between openness to experience and beliefs in both superstition and precognition. In contrast, no significant relationship between openness to experience and paranormal beliefs was found in Hitchman et al.'s (2012) study. Smith et al. (2009) suggest that a combination of openness to experience *and* sensation seeking personality traits is important for predicting paranormal beliefs. This is supported by previous work demonstrating a positive relationship between sensation seeking and paranormal beliefs (e.g., Pérez Navarro & Martínez Guerra, 2020; Kumar et al., 1993), though again there is some inconsistency in the reported significance of this relationship (see Swami et al., 2009; Tobacyk & Milford, 1983).

The most consistent findings reported in this area concern schizotypal personality traits and paranormal beliefs. The term "schizotypy" (introduced nearly 70 years ago by Rado, 1953) describes a broad phenotype of schizophrenic-like psychopathology and behavioural impairment (Mason & Claridge, 2015). Several models of schizotypy have been introduced which offer different definitions of the construct. Meehl's (1962) quasi-dimensional (or disease) model describes schizotypy as a milder form of schizophrenia, indicative of an increased risk of developing psychosis and a marker of psychological ill-health (French & Stone, 2014, pp. 56; Goulding 2004). Eysenck's (1960) model presents an opposing view, in which schizotypy can be described as a continuously variable personality dimension, with schizophrenia merely representing an extreme "end-point" of this continuum (Claridge & Beech, 1995, pp. 194). An alternative model offered by Claridge and colleagues (see Claridge, 1997; Claridge & Beech, 1995), built upon dimensional models of personality and psychopathology, describes schizotypy as a fully dimensional construct comprised of both pathological and healthy manifestations (Kwapil & Barrantes-Vidal, 2015). The fully dimensional model therefore argues that schizotypy acts within subclinical (traits part of normal individual differences expressed in the general population) and clinical (more extreme traits resulting in clinical disorders, i.e., schizophrenia) ranges (Kwapil & Barrantes-Vidal, 2015). In parallel to the literature seen for paranormal beliefs, research has demonstrated that increased schizotypal traits are associated with traumatic childhood experiences, including emotional, physical and sexual abuse (see Toutountzidis et al., 2022).

Empirical studies of paranormal beliefs in the general population support the idea of "healthy schizotypy" traits that differ from more clinical expressions. For example, Williams and Irwin (1991) found that while paranormal believers demonstrated higher schizotypy scores compared to control group participants (and higher paranormal belief scores that were comparable to schizophrenic and schizotypal groups), the profile of their schizotypal traits was different to the schizophrenic and schizotypal groups (e.g., lower expression of psychosis-related traits). More recent research also demonstrates differences in the specific dimensions of schizotypy that are associated with paranormal beliefs. Work by Dagnall et al. (2017) shows that dimensions of schizotypy relating to cognitive-perceptual deficits (e.g., magical thinking, unusual perceptual experience, and paranormal ideation) and disorganisation (odd behaviour and odd speech) demonstrate the strongest associations with paranormal beliefs. Darwin et al. (2011) found similar associations between paranormal beliefs and the cognitive-perceptual and disorganised dimensions of schizotypy, as well as interpersonal deficits (e.g., social anxiety, blunted affect, and a lack of close friends). In contrast, Goulding (2004) found no relationship between paranormal beliefs and aspects of schizotypy relating to cognitive disorganisation but did find a positive association with schizotypal traits relating to unusual experiences. While a full review of the literature concerning paranormal beliefs and schizotypy is beyond the scope of this work, the small number of findings reported here highlight the relationship between paranormal beliefs and both positive (dimensions relating to magical thinking and odd experiences) and negative schizotypal traits (dimensions relating to interpersonal deficits).

Cognition

The human cognitive systems are sophisticated collections of functions including memory, perception, reasoning, attention, language, and executive function (the umbrella term used to describe cognitive flexibility, inhibitory control, and working memory updating ability; for a full description see Chapter 4 of this thesis). Biases related to these cognitive functions, however, can systematically distort thinking in such a way that "can lead us to believe we have experienced events that did not really occur, or did not occur in quite the way in which we think we perceived them" (French & Stone, 2013, pp. 115; French & Wilson, 2007, pp. 3). In this respect, many cognitive biases have been related to paranormal beliefs and experiences. For example, confirmation bias (the tendency to seek out information that confirms one's beliefs and avoid information that challenges one's beliefs) has shown positive associations with paranormal beliefs in recent research (e.g., Lesaffre et al., 2021; Prike et al., 2018). It has been suggested, however, that those who are sceptical of the paranormal may also be susceptible to confirmation bias (for discussions on this topic, see Irwin et al., 2022). This is perhaps unsurprising when one considers that scepticism, or *dis*belief, in the paranormal is a type of belief in its own right (Lamont et al., 2009). Related to confirmation bias is subjective validation, a term used to refer to situations where an observer perceives two unrelated stimuli or events to correspond because they believe that they must (French & Stone, 2013; pp. 133). An example of subjective validation in relation to paranormal beliefs is the Barnum effect, named after American showman P.T. Barnum. The Barnum effect refers to the tendency for an individual to rate generic statements that could be applicable to many people as personally accurate and specific to themselves. It has been suggested that such statements are involved in the "cold reading" methods used by various individuals to make predictions about the past, present or future (see Hyman, 1981). Previous research has suggested that paranormal believers are more susceptible to the Barnum effect (Glick et al., 1989; Tobacyk et al., 1988), although this has not been extensively researched.

Previous studies have also documented a relationship between paranormal beliefs and thinking style (individual differences in the mechanisms and structures used in thinking and learning). Intuitive thinking has shown relatively robust positive associations with paranormal beliefs. For example, in a large sample of over 3,000 students from universities and vocational schools across Finland, Aarnio and Lindeman (2005) found positive associations between intuitive thinking and paranormal beliefs. This finding has not only been replicated by the authors in subsequent studies (e.g., Lindeman & Aarnio, 2006), but by other research groups (e.g., Williams et al., 2022; Genovese, 2005). With research demonstrating a positive relationship between paranormal beliefs and intuitive thinking, it follows that several studies have noted a negative relationship with analytical thinking styles (e.g., Tosyali & Aktas, 2021; Lindeman & Svedolm-Häkkinen, 2016; Svedholm & Lindeman, 2013). It should be noted, however, that many studies in this area rely on the use of undergraduate samples, and if paranormal beliefs were characterised by a complete lack of analytical thinking, paranormal beliefs would not be observable in these samples. In this respect, it might be more accurate to suggest that paranormal believers *prefer* intuitive thinking but can engage in analytical thinking styles when required.

A substantial volume of literature exists concerning relationships between paranormal beliefs and many cognitive functions not detailed here. To gain a clearer understanding of the role cognition plays in paranormal beliefs, a systematic review of the area would be required. This brief commentary, however, highlights how features of cognition (including biases and preferential ways of thinking) might contribute to the development and maintenance of paranormal beliefs.

Research questions

The present work takes a particular interest in the cognitions associated with paranormal beliefs and seeks to determine whether paranormal beliefs are associated with specific deficits in cognitive function. To this end, cognitive functions were examined using a mixture of self-report and behavioural methods to determine the direction and strength of their associations with paranormal beliefs.

The work will, therefore, consider the following research questions:

1. How reliable are current measures of paranormal beliefs? Specifically, is the widely used Revised Paranormal Beliefs Scale (RPBS; Tobacyk, 2004) a reliable measure of paranormal beliefs in the general population, or should (and can) a new, more reliable measure be developed?
2. Are paranormal beliefs associated with deficits in executive functioning? If so, are these deficits more prominent for specific executive functions?
3. Are there qualitative differences in believers' and sceptics' accounts of their own memories relating to their belief (or disbelief) in paranormal phenomena? If differences are found, how do these relate to existing theories of paranormal belief development?

CHAPTER 2. SYSTEMATIC REVIEW OF THE LITERATURE

INTRODUCTION

As discussed in the General Introduction chapter of this thesis, research on paranormal beliefs and cognition is extensive, spanning various aspects of cognitive functions. The last published review to examine the relationships between paranormal beliefs and cognition was conducted by Irwin (1993). That non-systematic narrative review of 43 studies is now almost 30 years old and may have introduced bias by "…citing null results only when these form a substantial proportion of the available data on a given relationship" (pp. 6). At the time of his review, Irwin (1993) concluded that, owing to the variable findings, support for the cognitive deficits hypothesis remained uncertain. Research has grown considerably since Irwin's (1993) review and an updated and systematic review is timely. The current review has two key aims: first, to provide the first assessment of study quality in this area and second, to systematically review and summarise key associations between paranormal beliefs and a range of cognitive functions.

METHOD

This review was conducted within the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The systematic review was preregistered at the Open Science Framework (OSF; https://osf.io/uzm5v) as part of a larger study. Data used for the descriptive and inferential analyses presented in the results section are available at the OSF preregistration.

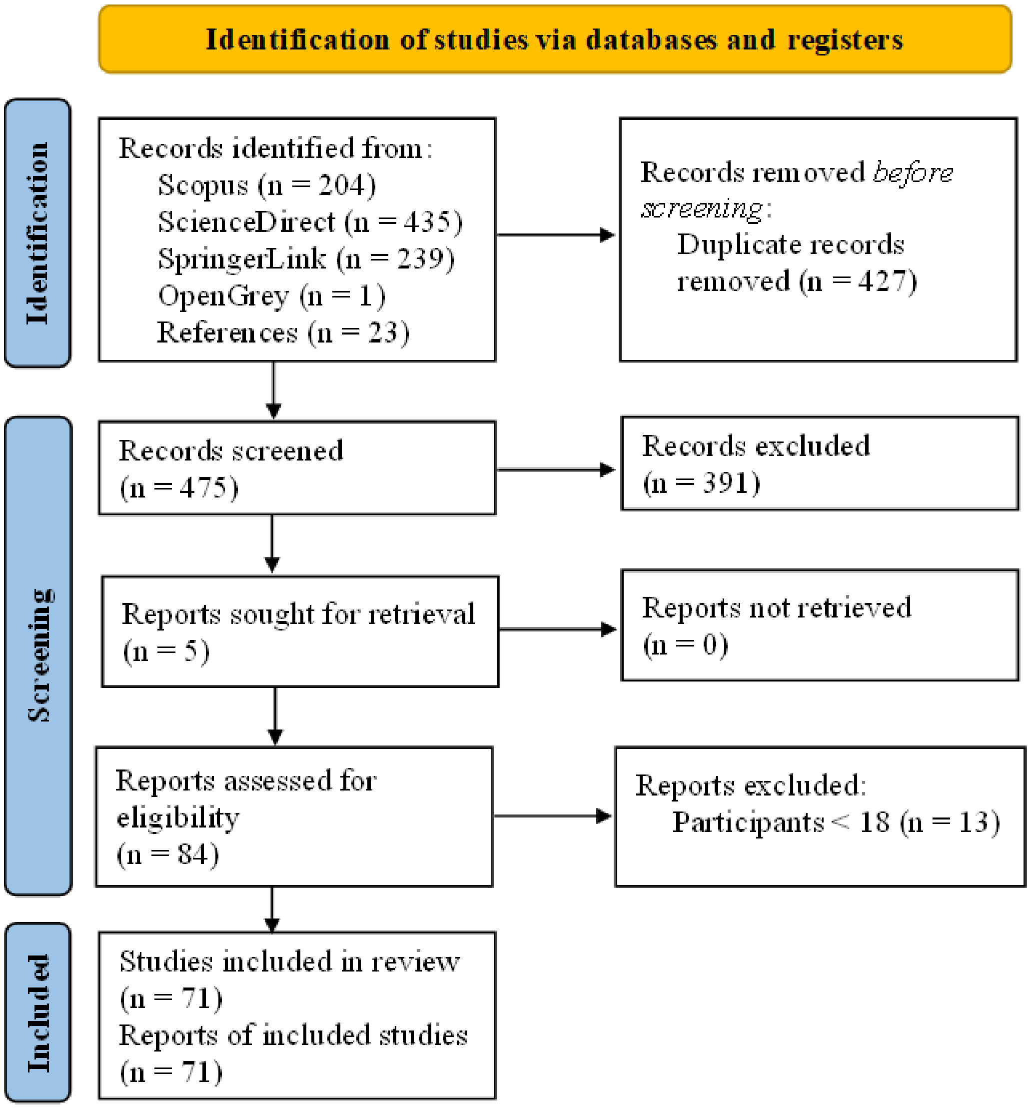
Search strategy

A systematic literature review was chosen for this area owing to its strength as a method to synthesise relevant evidence from large bodies of research (Mallett et al., 2012; Harari et al., 2020). Searches included both peer-reviewed articles published in scholarly journals and "grey literature" (concerning unpublished works such as doctoral theses). The electronic databases Scopus, ScienceDirect, SpringerLink, and OpenGrey were searched from inception to May 2021 using the following terms: (1) "paranormal belief" AND cogni*, (2) "paranormal belief" AND thinking, and (3) "paranormal belief" AND (memory OR "executive function"). For databases that did not permit wildcard Boolean operators (ScienceDirect), one of the above search terms was amended and entered as: "paranormal belief" AND (cognition OR cognitive), to best replicate the effect of the Boolean operator. Following exclusion of duplicate articles across databases, titles and abstracts were assessed to identify studies relevant to the review. Full-text assessment of eligible studies was performed to determine final inclusion. Full-text copies were unavailable for five studies, which were subsequently sought for retrieval. Finally, hand-searches of the reference lists for each included article were conducted to identify any additional relevant articles. The PRISMA flow diagram presented in Figure 2.1. illustrates the full screening and selection process.

Inclusion/exclusion criteria

Studies were eligible for inclusion if they were: published in the English language, conducted with a healthy adult sample (age 18 or over) and presented original data involving both a measure of paranormal belief and a measure of cognitive function. As cognitive functions have been shown to peak at different ages (for a detailed discussion on this topic, see Hartshorne & Germine, 2015), samples that included children and adolescents under the age of 18 were excluded from the review as some cognitive functions are still developing in these younger individuals.

Figure 2.1. PRISMA flow diagram.



Data extraction

A detailed data extraction form was used to collate the following information from included studies: sample sizes and demographic details (including sex, age and education), the measures of self-rated paranormal belief, the aspect of cognition assessed, the tests of cognitive functions used, and findings relating to the relationship between paranormal beliefs and cognitive function. Eligible outcome measures were broadly categorised to include both global cognitive function and domain-specific cognitive functions. Any measure of cognitive function was eligible for inclusion (e.g., neuropsychological tests, self-report measures). Results for both paranormal beliefs and cognitive functioning could be reported as an overall test score that provides a composite measure, subscale scores that provide domain-specific measures, or a combination of the two. When multiple cognitive outcomes were investigated, all measures were included. To assess the strength of the relationships between paranormal beliefs and various cognitive functions, the number of positive, negative, or null findings reported by each study included in the review were calculated. Measures of paranormal belief were examined to determine the extent to which established questionnaires have been used. In line with the preregistered protocol, evidence was synthesised narratively. Meta-analyses could not be undertaken because of the heterogeneity of study designs and outcome measures. Summary tables were developed, however, that include information relating to: sample size, gender composition, mean sample age, cognitive domain, outcome measure, and key findings. Given the range of outcome measures, categorisation of the included studies was attempted using common cognitive domains. As the review took an explorative approach, and did not specify domains of interest, categorisation took place after full-text evaluation of included studies.

RESULTS

Electronic and hand searches identified 902 papers, of which 475 were unique. Most articles (k = 391) were excluded from the review following title and abstract screening, leaving 84 eligible for full-text evaluation. Thirteen studies that included participants under the age of 18 were excluded from the review (see Appendix A for details of these studies). Seventy-one papers met the inclusion criteria (see Figure 2.1.), which included 70 published between 1980 and 2020 and one unpublished doctoral thesis (Greening, 2002).

Assessment of study quality and risk of bias

The preregistration for this review specified using a bespoke series of questions to assess study quality, but a more well-established and validated measure of study quality was subsequently used in the Appraisal tool for Cross-Sectional Studies (AXIS) tool (Downes et al., 2016). Of the 20 AXIS items, seven assess reporting quality (items: 1, 4, 10, 11, 12, 16 and 18), seven relate to study design (items: 2, 3, 5, 8, 17, 19 and 20), and six to possible biases (items: 6, 7, 9, 13, 14 and 15). Quality ratings were corroborated with an external researcher and these two sets of ratings had almost-perfect agreement (93%) with Kappa = .84.

Following previous research (Lannoy et al., 2021), AXIS quality scores were classified according to the number of "Yes" responses for the 20 items for each study—poor quality for scores for scores <50%, fair quality for scores between 50 to 69%, good quality for scores of 70% to 79%, strong quality for scores of 80% and higher. Three in four studies were rated as either 'strong' (26/71: 37%) or 'good' (27/71: 39%). By contrast, 17/71 (24%) were rated as 'fair' and only 1/71 (1%) was rated as 'poor'. The mean quality rating score across all 71 studies was in the 'good' range; however individual AXIS items are not weighted and so this total score provides a general, but limited, classification that should be interpreted with some caution. The number of papers meeting each AXIS criterion ('Yes') is presented in Table 2.1. The number of papers meeting the criteria for each AXIS domain (reporting quality, study design quality, and potential biases) is presented in Figures 2.2.-2.4., respectively.

All studies scored positively for items concerning: clear objectives, appropriate study design, appropriate measurement of outcome variables, internal consistency of presented results, and appropriate conclusions justified by the results. Study quality correlated with year of publication ($r = .64$, $p < .001$), and appears to be improving with time (see Figure 2.5.). Nonetheless, three main areas for study quality improvement were highlighted throughout the AXIS assessment: sample size justification, nonrespondents, and discussion of limitations.

Table 2.1. Total number of 'yes', 'no' and 'unsure' responses for each AXIS item.

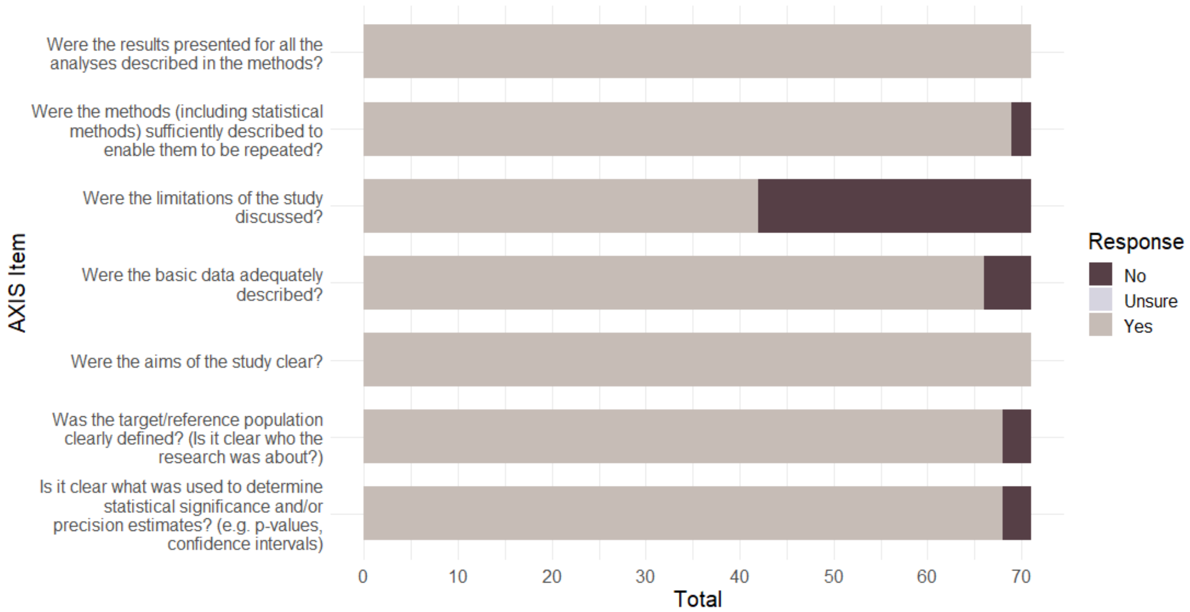
	AXIS Item	Yes	No	Unsure
Introduction				
1	Were the aims/objectives of the study clear?	71	0	0
Methods				
2	Was the study design appropriate for the stated aim(s)?	71	0	0
3	Was the sample size justified?	5	66	0
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)	68	3	0
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	22	49	0
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?	31	29	11
7	Were measures undertaken to address and categorise non-responders?	19	0	52
8	Were the risk factor and outcome variables measured appropriate to the aims of the study?	71	0	0
9	Were the risk factor and outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?	65	6	0
10	Is it clear what was used to determine statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)	68	3	0
11	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?	69	2	0
Results				
12	Were the basic data adequately described?	66	5	0
13	Does the response rate raise concerns about non-response bias?	7	12	52
14	If appropriate, was information about non-responders described?	1	18	52
15	Were the results internally consistent?	71	0	0
16	Were the results presented for all the analyses described in the methods?	71	0	0
Discussion				
17	Were the authors' discussions and conclusions justified by the results?	71	0	0
18	Were the limitations of the study discussed?	42	29	0
Other				
19	Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results?	0	14	57
20	Was ethical approval or consent of participants attained?	37	0	34

Sample size justification, representativeness, and open science

Only 5 of 71 (7%) papers included a-priori power analyses to justify their sample sizes. Although power analyses are rarely conducted in this research area, the mean sample size is large at 211 (median = 124), suggesting that both simple correlational and between-subject comparisons are well-powered to detect large (.99 and .98), moderate (.94 and .88) and potentially for small effect sizes (.72 and .72)–large, moderate and small effects being 0.7, 0.5 and 0.2 respectively (Cohen, 1988). Despite this, many studies have assessed multiple outcomes and/or multiple metrics derived from the same tests and so, a simple power analysis will mislead. As a rough metric on this issue, we calculated the number of p-values presented in the results section for each of the 71 papers. This revealed a mean number of p-values per study of 43 (median = 30) with a range from 1 (Brugger et al., 1991) to over 200 (Prike et al., 2018). So, despite relatively large samples, the possibility of type-1 errors remains high, especially when studies fail to adjust alpha levels for high levels of multiple testing. Only 12/71 studies employed some correction; eleven used a Bonferroni correction (Betsch et al., 2020; Andrews & Tyson, 2019; Denovan et al., 2018; Prike et al., 2018; Wilson, 2018; Irwin, 2015; van Elk, 2015; Irwin et al., 2014; Simmonds-Moore, 2014; Willard & Norenzayan, 2013; Schienle et al., 1996), and one used the Newman–Keuls adjustment (Krummenacher et al., 2010). Those studies that adjusted alpha levels tended to report more p-values than those that did not adjust (means 57 vs. 40). So, adjustment was made in fewer than one-in-five studies, most being published recently.

Despite good-strong quality ratings, some core features of open science practice including preregistration have yet to be embraced in this literature. Admittedly, we are assessing forty years of research and preregistration is a relatively recent innovation in psychology. Nonetheless, the Open Science Framework (OSF) began in 2013 as a repository for preregistrations–so potentially up to half of the 71 studies could have preregistered, yet only 2 (<3%) have done so (Ballová Mikušková & Čavojova, 2020; Betsch et al.2020), with both published in 2020. The issue about reregistration is fundamental in this area of research. First, studies are characterised by large numbers of analyses often involving multiple outcome measures and/or multiple metrics derived from smaller numbers of tests. Up to one-third of studies (25/71) have assessed relationships between cognitive function and paranormal test subscale scores (often with few items), an approach which consciously or unconsciously increases the likelihood of reporting bias and HARKing (hypothesising after results are known), often perhaps with little chance of, or interest in, replicating such findings (see Laws, 2013, for a discussion). Second, the preregistration of future trials will also help to assess whether null results remain unpublished. Third, preregistration would identify both the primary outcome and the sample size required to achieve an acceptable level of statistical power. Ironically, the lack of attention to pre-registration and justifying sample sizes contrasts with research on paranormal phenomena, where study registration and a priori power calculations have been employed for many years (Cardeña, 2018).

Figure 2.2. AXIS reporting quality summary for the 71 papers included in the review.

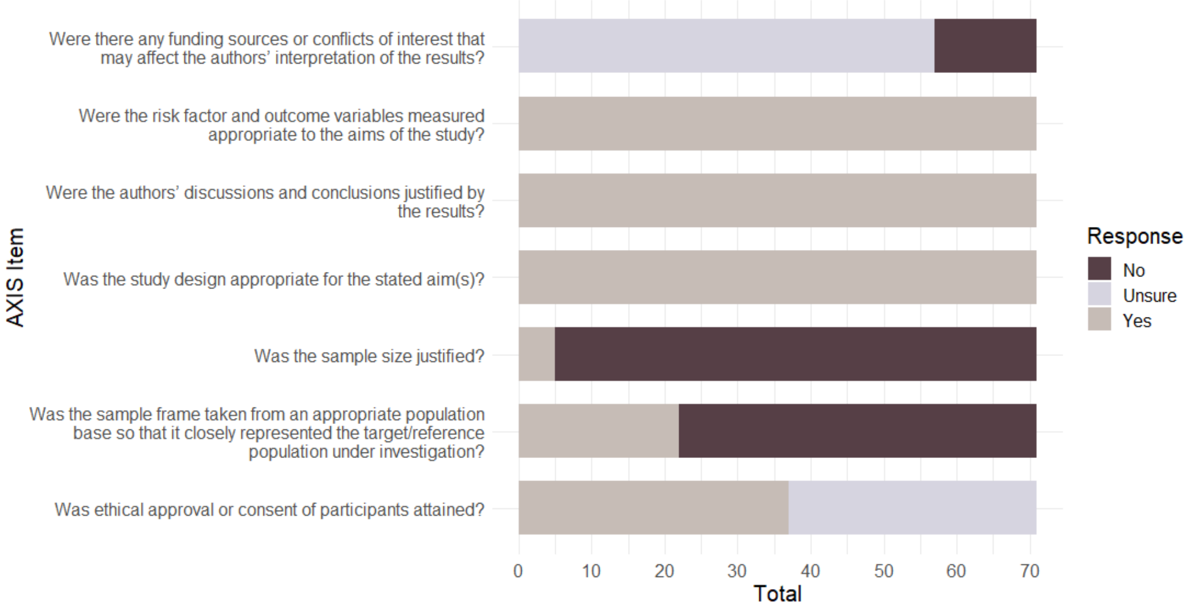


Representativeness

Another issue concerns the sampling frame and its representativeness. Almost two-thirds of all samples are undergraduates (45/71: 63%) and of those, 21 (30%) consisted wholly of, or a majority of, psychology undergraduates. Only one-third of all samples consisted of: non-undergraduates (15/71: 21%), mixed undergraduate and general population samples (8/71: 11%) or other non-undergraduate samples (2/71: 3%). One non-undergraduate study by Blackmore (1997) consisted of a national newspaper-based study (Daily Telegraph) and recruited an exceptionally large sample (n = 6238). If this outlier is excluded, then 60% of all participants in the 70 remaining studies have been completely (k = 41) or majority undergraduate (k = 5) samples, with 16 involving only psychology graduates. Amongst the non-undergraduate samples, this includes visitors to a paranormal fair (van Elk, 2017; van Elk, 2015), members of the Society for Psychical Research (Lawrence & Peters, 2004), Mechanical Turk participants (Gray & Gallo, 2016), and some used Crowdflower, a crowdsourcing website (Prike et al., 2018; Ståhl & Van Prooijen, 2018; Prike et al., 2017). So, even the non-undergraduate samples may not necessarily represent the wider population (see Stroebe et al., 2018, for a discussion). Studies testing undergraduates and non-undergraduates did not differ in mean sample size (196 vs 215, with the exclusion of Blackmore, 1997, $t(68) = .29$, $p = .78$, $d = .08$) or in quality ratings (14.73 vs 15.19: $t(69) = -.90$, $p = .37$: $d = .23$). The profile of sampling, however, is pertinent because paranormal beliefs are inversely related to educational levels (Bader et al., 2012; Van den Bulck & Custers, 2010; Sparks & Miller, 2001), and those studying sciences, medicine, and psychology exhibit lower levels of paranormal beliefs (Andrews & Tyson, 2019; Aarnio & Lindeman, 2005). Such samples are unrepresentative and may bias findings because they may combine lower levels of paranormal beliefs and higher cognitive functioning than occurs in the general population.

In addition to samples comprising more highly educated university students, most participants are female (>60%). The importance of this latter aspect of sampling is underscored for at least two reasons. First, some authors have documented greater levels of paranormal beliefs in women (Ward & King, 2020; Rogers et al., 2018; van Elk, 2017; Rogers et al., 2016; Voracek, 2009; Watt et al., 2007). Indeed, the last literature review by Irwin (1993) stated that "the endorsement of most, but certainly not all, paranormal beliefs is stronger among women than among men" (pp.8). Second, gender (and age) effects are not consistently reported (Lange et al., 2001) and have resulted in substantial debate (Irwin, 2000; Vitulli, 2000; Vitulli et al., 1999). This debate largely results from differences in psychological test theories (see Dean et al., 2021, for a discussion). Classical test theory—used to develop common paranormal belief measures, such as the RPBS—does not test for the presence of differential item functioning (DIF). DIF refers to when individuals with the same latent ability (e.g., paranormal beliefs), but from different groups, have an unequal probability of giving a response. By contrast, modern test theory, including the use of Rasch scaling, can produce unbiased interval measures focused on the hierarchical properties of questionnaire items. This has resulted in the revision of older paranormal belief measures using modern test theory, to create scales that accurately capture fluctuations in levels of belief rather than differences in item functioning (Lange & Thalbourne, 2002; Lange et al., 2001). When these problematic items are removed from scales such as the RPBS and ASGS, paranormal belief scores are no longer associated with sex, but small differences remain for age (Lange & Thalbourne, 2002; Lange et al., 2001). Although these effect sizes seem to be small (e.g., 0.15; Lange & Thalbourne, 2002, identified by Cohen, 1988, as a small effect size), they are more likely to reflect a true and meaningful fluctuation in paranormal belief levels, compared to findings reported using scales developed through classical test theory.

Figure 2.3. AXIS study design quality summary for the 71 papers included in the review.

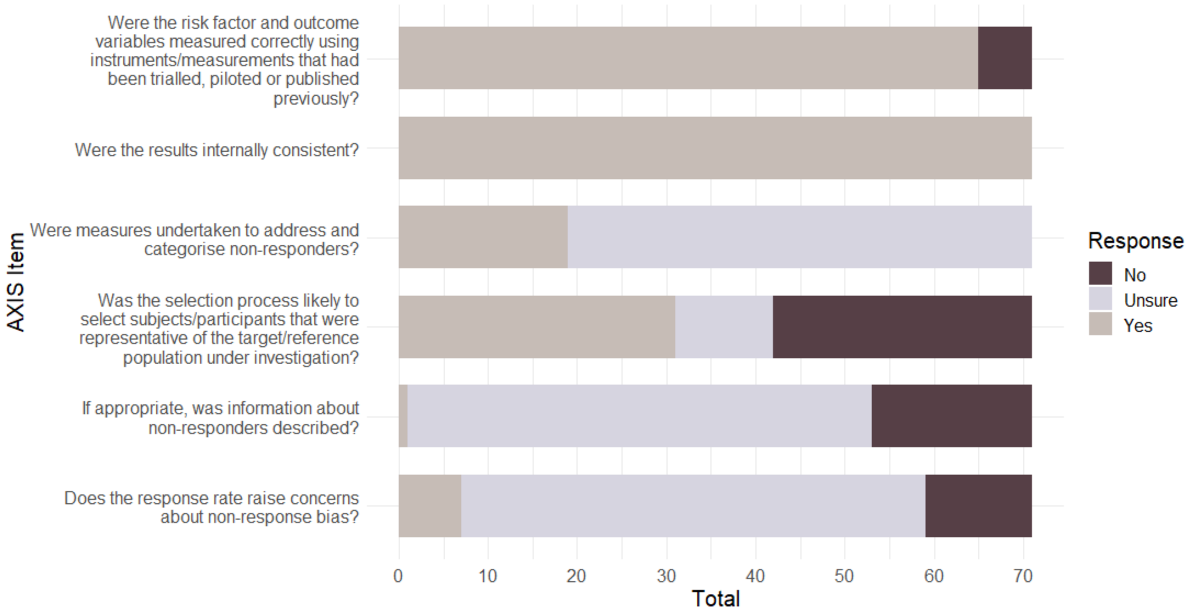


Nonrespondents

Most studies (52/71) failed to state whether measures were undertaken to address and categorise nonrespondents. As such, response rates and risk of nonresponse bias could not be calculated. Nonresponse bias arises when respondents differ from nonrespondents beyond sampling error and may reduce external validity (Werner et al., 2007; Hawkins, 1975). Survey-based approaches are at a greater risk of nonresponse bias owing to their high nonresponse rates, with those relying on self-administered online surveys suffering from higher nonresponse rates than those using face-to-face methods (Tobacyk, 1983). Most studies have been conducted in face-to-face settings (k = 59), however the past few years has seen a rise in online data capture (k = 12). Compared to face-to-face studies, online studies rated more highly on study quality (16.50 vs 14.49: $t(69) = -3.87$, $p < .001$, $d = 1.32$) and had larger mean sample sizes (482 vs 155: $t(11.83) = -3.12$, $p = .008$, $d = -1.69$, equal variances not assumed), but also report larger numbers of statistical comparisons (96.42 vs 31.58,: $t(12) = -3.47$, $p = .005$, $d = 1.33$, equal variances not assumed).

Of the 19 papers that did provide nonresponse rates, seven had response rates < 70% and so raise concerns about potential nonresponse bias (Prince, 2012). Only one of 19 papers (Lindeman & Svedholm-Häkkinen, 2016) presented any information about nonrespondents, reporting that they had marginally lower educational attainment than respondents. Similar findings for nonrespondents have been reported in other research areas (Kontto et al., 2020; Tolonen et al., 2005; Dalecki et al., 1993; Gannon et al., 1971). Finally, it should be noted that online studies more often have records of nonrespondents. Guidance has been developed on detailing non-response details in online survey-type studies e.g., the Checklist for Reporting Results of Internet E-Surveys (CHERRIES; Eysenbach, 2004) and should routinely be reported.

Figure 1.4. AXIS possible biases summary for the 71 papers included in the review.



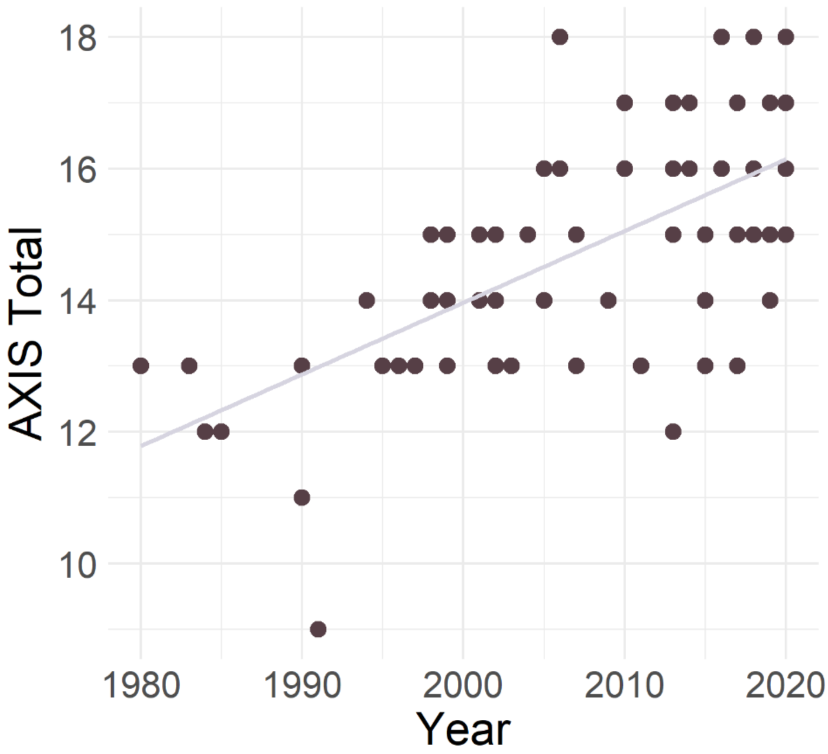
Limitations

Surprisingly, up to 40% of the included papers (29 of 71) did not include a discussion of study limitations. Discussion of study limitations forms a fundamental part of scientific discourse and is crucial for genuine scientific progress, allowing a reader to contextualise research findings (Ioannidis, 2007). The failure to discuss limitations might be viewed partly as a failure of the peer review process (Horton, 2002), but responsibility ultimately resides with authors. Detailing limitations allows other researchers to consider methodological improvements, identify gaps in the literature and has an ethical element by aiding research transparency. The inclusion of limitations not only helps increase research quality, but facilitates directions for future research and crucially, replications.

Quality summary

Of the 71 studies published since 1980, three-quarters were rated as 'good' or 'strong' in quality, and only one received a 'poor' quality rating. Indeed, study quality also indicates a continuous improvement in study quality across four decades of research. Despite the high levels of study quality and evidence of improving quality, several areas of methodological weakness were identified: justifying sample size, providing more detail about non-respondents, and discussing study limitations. One issue of note is the sampling, where almost two in three studies have relied on exclusively undergraduate samples (46/71: 65%), with many being psychology undergraduates. Future recruitment needs to move beyond the highly educated and address the bias towards female participants. Despite recruiting large samples, studies use large numbers of analyses, with a mean of 43 *p*-values reported in results sections, and rarely report appropriate adjustment of significance levels (12/71: 17%). These methodological issues are compounded by the fact that so few studies pre-register their primary hypotheses and analyses in advance (2/71: 3%).

Figure 2.5. AXIS study quality (maximum = 20) by year of publication.



Cognitive functioning

The 71 studies were grouped into six sections: (1) perceptual and cognitive biases, (2) reasoning, (3) intelligence, critical thinking, and academic performance, (4) thinking style, (5) executive function, and (6) other cognitive functions. Whenever possible, categories were classified according to the focus identified by the authors in each study. Such classifications are necessarily a simplification and not intended to provide a definitive organisation. Moreover, many studies could receive multiple classifications owing to the breadth of testing conducted (see Appendix B). In this context, Appendix B shows that two in three (48/71) studies might be classified as assessing executive function.

Articles presented in the first section (perceptual and cognitive biases) included scenarios aimed at measuring cognitive biases towards confirmatory evidence, and the impact of visually degraded stimuli on biases in perceptual decision-making. Examples of tasks used in the second section (reasoning) include the mental dice task (Brugger et al.,

1991) aimed at measuring probabilistic reasoning, and the Reasoning Tasks Questionnaire (RTQ; Blackmore & Trościanko, 1985) to assess both probabilistic and conditional reasoning. Studies in the third category (intelligence, critical thinking, and academic performance) included published measures such as the Watson-Glaser Critical Thinking Appraisal (WGCTA; Watson, 1980) and variations of Raven's matrices (e.g., the Advanced Progressive Matrices Test; Raven, 2000, Raven's Progressive Matrices; Raven et al., 2000, and measures of academic achievement such as grade point average). In the fourth section (thinking style), papers used measures such as the Rational Experiential Inventory (REI; Pacini & Epstein, 1999) and the Cognitive Reflection Test (CRT; Frederick, 2005), aimed at assessing intuitive and analytical thinking. Studies in the fifth section (executive function and memory) included tasks such as the Deese-Roediger-McDermott task (DRM; Roediger & McDermott, 1995) and the Wisconsin Card Sorting Test (WCST; Berg, 1948; Grant & Berg, 1948). The final cognitive section (other cognitive functions) included tasks to measure indirect semantic priming (using prime-target word pairs) and implicit sequence learning. Summary statistics for studies in each section can be found in Appendix C-H.

Perceptual and cognitive biases

Nineteen articles (n = 3,397) assessed perceptual and cognitive biases. Perceptual decisionmaking with high visual noise stimuli has produced inconsistent findings (k = 7). For example, in Simmonds-Moore (2014) found believers made more misidentifications of degraded black and white images of objects and animals (e.g., shark, umbrella), despite having faster response latencies than sceptics (suggesting a potential speed-error trade-off, with believers favouring speed over accuracy). By contrast, Van Elk (2015) found sceptics mis-categorised degraded black and white images of face stimuli as houses more frequently than believers. The findings from both studies, however, contradict those from Blackmore and Moore's (1994) study, which reported no difference in the accurate identification of degraded monochrome images for believers and sceptics. Two studies assessed perceptual decision-making relating to faces within degraded and artifact stimuli. Using black and grey images of faces and "nonfaces" (scrambled eyes-nose-mouth configurations), Krummenacher et al. (2010) found believers made significantly more Type I errors than sceptics, favouring "false alarms" over "misses" (i.e., believers had a lower response criterion when classifying images as faces, with a bias towards "yes" responses). Similarly, Riekki et al. (2013) presented participants with 98 artifact face pictures (containing a facelike area where eyes and a mouth could be perceived, e.g., a tree trunk) and 87 theme-matched non-face pictures (e.g., a tree trunk with no face-like areas). Believers rated the non-face pictures as more face-like and assigned more extreme positive and negative emotions to non-faces than sceptics. A study conducted by Caputo (2017) employed the strange-face illusion paradigm, in which pairs of participants are instructed to gaze into each other's eyes for 10 minutes in a dimly-lit room. This paradigm induces the experience of seeing face-related illusions and is assessed on a self-report measure (Strange Face Questionnaire, SFQ; Caputo, 2015). No association was found for paranormal beliefs and the experience of strange-face illusions. A final study of perceptual decision-making conducted by Van Elk (2013) used point-light-walker displays (an animated point-set of 12 points, representing a human walking on a treadmill), randomly scrambling the location of each individual dot across the display; and participants had to detect if a human agent was present. Paranormal believers were more prone to illusory agency detection than sceptics, being biased towards 'yes' responses when no agent was present. Cognitive biases have been assessed in 11 papers. These include reports of significant associations between paranormal belief and illusion of control or differences in causation judgements (Griffiths et al., 2019; Blanco et al., 2015; Rudski, 2004; Schienle et al., 1996) and risk perception (Drinkwater et al., 2019). Two studies, however, report no significant relationships (van Elk, 2017; Gagné & McKelvie, 1990). Further work shows that paranormal beliefs positively correlated with biases towards: anthropomorphism, dualism, teleology, and mentalising, but were not predicted by mentalising (Willard & Norenzayan, 2013). Proneness to jump to conclusions was assessed by Irwin et al. (2014) using a computerised task (see Dudley et al., 1997). Participants were informed of proportions of beads in two jars (e.g., 70 black and 30 red beads in jar one, but 30 black and 70 red beads in jar two), then shown a sequence of beads drawn one at a time from one of the jars and asked to identify whether beads were drawn from jar one or two, and to indicate when they are certain. Those who require fewer draws before being certain of their decision are identified as being prone to "jump to conclusions". A significant negative correlation emerged for jumping to conclusions, but only with the Traditional Religious Beliefs (TRB) subscale of the Rasch-devised RPBS (see Lange et al., 2000). A significant positive correlation was also found between TRB scores and self-report indices of jumping to conclusions as measured with the Cognitive Biases Questionnaire (Lesaffre et al., 2020; Peters et al., 2004; e.g., "imagine you hear that a friend is having a party and you have not been invited", 1 = little or no inclination to jump to a premature conclusion, 2 = inclination to make a cautious inference, 3 = inclination to jump to a dramatic inference). Prike et al. (2018) assessed proneness to jumping to conclusions using both a neutral (beads task) and an emotional draws-to-decision task (where participants decide whether positive or negative words are more likely a description of "Person A" or "Person B"—for a full description see Dudley et al., 1997). Participants also saw a series of 24 scenarios to assess bias towards confirmatory and disconfirmatory evidence, as well as liberal acceptance. Each scenario consisted of three statements presented one at a time, e.g., (a) "Eric often carries binoculars with him", (b) "Eric always has an unpredictable schedule", (c) "Eric tries to solve mysteries". Participants rated the likelihood of the same four response options after each statement, e.g., (a) "Eric is a private detective", (b) "Eric is a bird expert", (c) "Eric is a stalker", (d) "Eric is an astronaut". Each scenario presented an absurd interpretation (implausible for all three statements), a neutral lure, an emotional lure, and a true interpretation (less or equally as plausible as the lure options after the first statement but became the most plausible by the third statement). Paranormal beliefs were related to both disconfirmitory and confirmatory biases, but not to jumping-to-conclusions. Liberal acceptance predicted belief in the paranormal, but not after controlling for delusion proneness (as measured by the Peters et al. Delusions Inventory, PDI; Peters et al., 2004). Lesaffre et al. (2020) exposed participants to a magic performance and asked whether it was accomplished through: (1) paranormal, psychic, or supernatural powers, (2) ordinary magic trickery, or (3) religious miracles. Confirmation bias (i.e., explaining the magic performance in terms of paranormal powers) was associated with higher levels of paranormal beliefs. Barberia et al. (2018) demonstrated that educating participants about confirmatory bias reduced scores on the Precognition subscale of the RPBS (but did not reduce global belief scores).

Summary

The studies assessing perceptual and cognitive biases are somewhat inconsistent regarding perceptual decision-making errors in response to degraded or ambiguous stimuli. Of the studies exploring perceptual decision-making, four suggest an inverse relationship between paranormal belief and perceptual decision-making, two found no relationship, and one reported more perceptual decision-making errors from sceptics. Results show greater consistency when perceptual decision-making tasks involve identifying a human face/agent (rather than inanimate objects or animals), with believers making significantly more false-positive misidentifications than sceptics. In the 11 studies exploring cognitive biases, paranormal believers show a consistent bias towards both confirmatory and disconfirmatory evidence. The evidence that paranormal belief links to the tendency to "jump to conclusions" is weaker, but only two studies present findings related to this outcome.

Reasoning

Seventeen papers have focussed on reasoning ability (n = 9,661), with the majority (12/17) reporting significant inverse relationships with paranormal beliefs and probabilistic reasoning. Perception of randomness and the conjunction fallacy have also been associated with paranormal beliefs on tasks with both neutral and paranormal content (Denovan et al., 2018; Prike et al., 2017; Dagnall et al., 2016a; Dagnall et al., 2016b; Dagnall et al., 2014; Rogers et al., 2009; Dagnall et al., 2007). Dagnall et al.'s (2007) study presented 17 reasoning problems across four categories: perception of randomness, base rate, conjunction fallacy, and probability. Perception of randomness problems required participants to determine the likelihood of obtaining particular strings (e.g., "Imagine a coin was tossed six times. Which pattern of results do you think is most likely? (a) HHHHHH, (b) HHHTTT, (c) HTHHTT, (d) all equally likely"). Performance on these problems significantly predicted paranormal belief, with believers making more errors than sceptics. No significant differences or predictive effects emerged for the three other problem categories. In a later study, Dagnall et al. (2014) presented 20 reasoning problems across five categories of: perception of randomness, base rate, conjunction fallacy, paranormal conjunction fallacy, and probability. The authors again reported perception of randomness to be the sole predictor of paranormal beliefs, with high belief associated with fewer correct responses. While these papers report no effects in relation to conjunction fallacy, Rogers et al. (2009) demonstrated a significant main effect of paranormal belief on conjunction errors, with believers making more errors than sceptics. In later studies, both Prike et al. (2017) and Rogers et al. (2016) reported an association between paranormal belief and conjunction fallacy, but this association was only significant for scenarios with confirmatory outcomes in the latter study. Probabilistic reasoning ability has been consistently associated with paranormal beliefs across five studies. In one paper (Musch & Ehrenberg, 2002), participants received a probabilistic reasoning test battery comprised of six tasks. For example, one task was a variant of the birthday paradox (from Blackmore & Trościanko, 1985), in which participants are asked: "How many people would you need to have at a party to have a 50:50 chance that two of them will have the same birthday (regardless of year of birth)". Possible answers for this task were 22 (correct), 43, or 98. Significant positive correlations emerged between paranormal beliefs and errors on three of the six tasks (dice sequences, dice throws, and sample size estimates). In the second study (Brugger et al., 1991), participants received written descriptions of two hypothetical events: throwing 10 dice once to get 10 sixes and throwing one die 10 times to get 10 successive sixes; and had to identify whether one event was more probable or both equally probable. The authors reported 64% of believers and 80% of sceptics correctly identified that both events were equally probable. Brugger et al. (1990) assessed differences in repetition avoidance between believers and sceptics on a mental dice task (where participants imagined throwing a die and had to write down the number they imagined being on top of the die), finding significantly fewer repetitions in believers than sceptics. Similarly, Bressan (2002) used a probabilistic reasoning questionnaire with problems concerning the comprehension of sampling issues, sensitivity to sample size, representative bias (as applied to sample size or random sequences) and the generation of random sequences. Believers made more probabilistic errors on two of four generation of random sequences problems: (1) simulated coin toss problem, in which participants were asked to fill in 66 empty cells by writing 'H' (heads) or 'T' (tails) randomly to make a resulting sequence that was indistinguishable from that of an actually tossed coin), and (2) an adapted version of Brugger et al.'s (1990) mental dice task. Finally, Blackmore (1997) asked participants whether a list of 10 statements (as might be produced by a psychic, e.g., "there is someone called Jack in my family") were true for them, and to estimate the number of these statements that might be true for a stranger in the street. The number of 'true' statements was greater for believers than sceptics (significantly on five of the ten questions), however no significant differences emerged when estimating the number of statements true for a stranger. The final four papers in this section found non-significant correlations between paranormal belief and probabilistic reasoning, but significant correlations with conditional reasoning tasks. Using the Reasoning Tasks Questionnaire (RTQ; see Blackmore & Trościanko, 1985), one study (Pérez Navarro & Martínez Guerra, 2020) found neither probabilistic reasoning nor neutral conditional reasoning were associated with paranormal beliefs. However, conditional reasoning was associated with paranormal beliefs when conditional reasoning tasks contained paranormal content rather than neutral content, with believers making fewer errors on these tasks. The second paper (Roberts & Seager, 1999) measured reasoning using a test that combined probabilistic reasoning questions (seven in total, four of which were derived from the RTQ), conditional reasoning questions with abstract content (e.g., "if C is true, then D will be observed. D is observed. Therefore, C is true: True or False?"), and conditional reasoning questions with paranormal content (e.g., "if people are aware of hidden objects, then clairvoyance exists. People are aware of hidden objects. Therefore, clairvoyance does exist: True or False?"). Overall, paranormal beliefs correlated negatively with reasoning ability and conditional reasoning ability, but not with probabilistic reasoning ability. When comparing the two types of conditional reasoning questions, the authors reported no difference between the correlations for paranormal beliefs and either the abstract or paranormal conditions. Following a similar format, Wierzbicki (1985) assessed reasoning ability using 16 conditional reasoning statements with either parapsychological or abstract content, finding paranormal belief scores and number of reasoning errors correlated positively. The final paper in this section (Lawrence & Peters, 2004) employed 32 statements conditional reasoning statements and found participants with strong paranormal beliefs made more reasoning errors than those with weak paranormal beliefs.

Summary

In general, evidence suggests paranormal beliefs are associated with poorer reasoning, however this line of research is characterised by inconsistent findings. Two studies report that the perception of randomness is a significant predictor of paranormal belief and provide some evidence of replicability (Dagnall et al., 2014; Dagnall et al., 2007). Despite this, evidence regarding the association between paranormal belief and the conjunction fallacy are conflicting, with two studies (Dagnall et al., 2014; Dagnall et al., 2007) reporting no effect, and three (Prike et al., 2017; Rogers et al., 2016; Rogers et al., 2009) reporting significant associations. This may be due, in part, to the different statistical techniques used within each study, as those reporting no effect (Dagnall et al., 2014; Dagnall et al., 2007) used multiple regression analyses with all probabilistic tasks entered as predictor variables, while studies reporting significant associations (Prike et al., 2017; Rogers et al., 2016; Rogers et al., 2009) only included conjunction fallacy tasks in their predictive models. Similar inconsistency emerges for probabilistic reasoning, with nearly equal numbers of studies reporting significant and nonsignificant associations with paranormal beliefs.

Intelligence, critical thinking, and academic performance

Twelve studies explored intelligence, critical thinking, and academic performance ($n = 2,657$). Seven papers focused on critical thinking ability, with two finding significant reductions in paranormal belief following a course in critical thinking (Wilson, 2018; McLean & Miller, 2010). Alcock and Otis' (1980) study employed the Watson-Glaser Critical Thinking Appraisal (WGCTA; see Watson & Glaser, 1964) significantly higher levels of critical thinking ability in sceptics than believers. Morgan and Morgan (1998) conducted a similar study, measuring critical thinking using a revised version of the WGCTA (see Watson, 1980), finding significant negative correlations between critical thinking ability and three subscales of the PBS (Superstition, Traditional Religious Belief, and Spiritualism). No significant correlation between paranormal belief and critical thinking emerged in the remaining three papers (Hergovich & Arendasy, 2005; Roe, 1999; Royalty, 1995). One did, however, report significant negative correlations between reasoning ability (measured using the Winer Matrizen-Test, see Formann & Piswanger, 1979) and three subscales of the PBS: Traditional Paranormal Beliefs, Traditional Religiosity, and Superstition (Hergovich & Arendasy, 2005). The links between paranormal beliefs and academic achievement, or general intelligence are both mixed and weak. Two papers report significant negative correlations, one between overall paranormal belief scores and mean academic grade (Andrews & Tyson, 2019) and one between grade point average and the Witchcraft and Superstition subscales of the PBS (Tobacyk, 1984). Turning to intelligence, Betsch et al. (2020) found a significant inverse relationship between IQ and paranormal beliefs, but only when controlling for sex, supporting similar findings from Smith et al.'s (1998) study which reported a significant negative correlation between paranormal beliefs and intelligence (using the Advanced Progressive Matrices Test, Set 1, see Raven, 1976). Nevertheless, two studies found no association between paranormal beliefs and intelligence. Royalty (1995) used the information subtest of the Wechsler Adult Intelligence Scale (Wechsler, 1955) as an estimate of fullscale IQ, and the vocabulary subtest of the Multidimensional Aptitude Battery (Jackson, 1985) as a measure of verbal intelligence. Stuart-Hamilton et al. (2006) found no relationship with fluid intelligence using Raven's Progressive Matrices (Raven et al., 2000); however, this sample were older (mean age of 71).

Summary

Conflicting findings emerge from studies of intelligence, critical thinking, and academic performance, with an almost equal number of significant and non-significant associations to paranormal beliefs. Some of this heterogeneity, however, appears to reflect whether studies used crystallised or fluid intelligence tasks and the age of the sample (e.g., Stuart-Hamilton et al. (2006) failed to find a relationship between fluid IQ and paranormal beliefs in an older sample, but Smith et al. (1998) found a significant negative association in a younger sample). The precise relationship of paranormal belief with intelligence requires further investigation, both by considering the age of the sample and assessing relationships with fluid and crystallised intelligence separately.

Thinking style

Thirteen studies (n = 4,100) examined aspects of thinking style. One consistent finding is a significant association between paranormal belief and an intuitive thinking style, which is characterised as being quick and guided by emotion (Branković, 2019; Lasikiewicz, 2016; Majima, 2015; Svedholm & Lindeman, 2013; Genovese, 2005). A further study (Rogers et al., 2019) also reports a significant partial correlation after controlling for sample type (online versus recruited face-to-face recruitment) owing to significantly higher levels of paranormal beliefs and intuitive thinking, and significantly lower rational/analytical thinking, in the online sample versus the face-to-face sample. Contradictory findings, however, have emerged concerning paranormal beliefs and an analytical thinking style, which is thought to be more effortful and driven by logic. A positive relationship emerged in two studies (Lasikiewicz, 2016; Majima, 2015) while two (Irwin, 2015; Genovese, 2005) found no relationship between paranormal beliefs and analytical thinking as assessed by the Rational Experiential Inventory (REI; see Pacini & Epstein, 1999). Four further studies report significant negative relationships between paranormal beliefs and analytical thinking using various measures: two (Ståhl & Van Prooijen, 2018; Rizeq et al., 2021) used different versions of the Cognitive Reflection Test (CRT; see Frederick, 2005); one (Lindeman & Svedholm-Häkkinen, 2016) used the Rational Experiential Multimodal Inventory (see Norris & Epstein, 2011); and one (Rogers et al., 2019) used both the Argument Evaluation Test (AET; see Stanovich, 1997) and the Actively Open-Minded Thinking scale (AOT; see Sá et al., 1999; Stanovich, 1997). A further study reported a significant negative relationship between paranormal beliefs and analytical thinking but could not replicate the finding (Ballová Mikušková & Čavojová, 2020). The final two papers in this section document relationships between paranormal belief and other cognitive styles. Gianotti et al. (2001) presented participants with 80 word-pairs (40 semantically indirectly related, 40 semantically unrelated), and they had to state if a third noun was semantically related to both words. Believers showed increased verbal creativity, making significantly more rare associations than sceptics for unrelated word-pairs, but not for indirectly related word-pairs. Hergovich (2003) used the Gestaltwahrnehmungstest (see Hergovich & Hörndler, 1994) to assess degree of field dependence, by presenting participants with figures in which they needed to find an embedded figure in the form of a house and reported a significant positive relationship between paranormal beliefs and field dependence.

Summary

Eight papers report positive associations between an intuitive thinking style and paranormal belief (although it should be noted that one study reported only a partial correlation after controlling for sample type). By contrast, evidence concerning an analytical thinking style is inconsistent, with reports of a negative relationship with belief ($k = 4$), a positive relationship ($k = 2$), and no relationship ($k = 2$). An additional study did report a negative relationship between analytical thinking and paranormal belief, but this was not replicated in a follow-up study. The final two studies in this section suggest positive relationships between paranormal belief and both verbal creativity and field dependence.

Executive function and memory

Six studies ($n = 810$) assessed memory or executive function. Turning first to memory, the findings are inconsistent. One study (Wilson & French, 2006) showed paranormal belief predicted false memory responses on a questionnaire-based measure, and two others

(Lawrence & Peters, 2004; Greening, 2002) reported associations between belief and behavioural measures of false memories but failed to replicate this in additional samples. Dudley's (1999) study had participants complete the Paranormal Belief Scale while rehearsing a five-digit number or not; and found significantly higher paranormal belief scores in the group who had their working memory restricted (by the rehearsal task). However, a recent study by Gray and Gallo (2016) failed to find any differences in working memory, episodic memory or autobiographical memory for believers and sceptics. Further inconsistencies can be seen when exploring relationships between paranormal belief and inhibitory control, with Lindeman et al. (2011) noting more errors from believers than sceptics on the Wisconsin Card Sorting Test (WCST; see Berg, 1948; Grant & Berg, 1948), but not on the Stroop task (see Stroop, 1935). Wain and Spinella (2007) explored executive function using a self-report measure and found a negative correlation between paranormal belief and executive functioning, with negative correlations between belief and both inhibition and organisation.

Summary

The studies in this section report inconsistent links between paranormal belief and memory. While three of four memory studies report links between paranormal beliefs and an increased tendency to create false memories, two of these studies failed to replicate the finding. Two studies assessing executive functioning both suggest poorer performance is associated with belief but may interact with the measure of executive functioning.

Other cognitive functions

Finally, four papers (n = 368) explored other aspects of cognitive function not covered by the categories already described. Pizzagalli et al. (2001) tested the association between indirect semantic priming and paranormal beliefs using 240 prime-target word pairs, with target words either directly related, indirectly related, or unrelated to the prime word.

Compared to sceptics, believers had shorter reaction times for indirectly related target words were presented in the left visual field, suggesting a faster appreciation of distant semantic associations which the authors view as evidence of disordered thought. The final three papers did not find any significant relationships between paranormal beliefs and: implicit sequence learning (Palmer et al., 2007), cognitive complexity (Tobacyk, 1983), or central monitoring efficiency (Irwin & Green, 1999).

Critical evaluation

A critical evaluation of this work highlights a prominent issue within the research area; namely that study findings across the domains outlined above are largely inconsistent. This is perhaps unsurprising when we consider that psychology as a discipline is facing a replication crisis. With reproducibility being an essential part of the scientific method, failures to replicate previous research findings undermine the credibility of theories and scientific knowledge. As such, multiple replication studies will often be needed to resolve inconsistent findings in the literature (Maxwell et al., 2015). The value of such a standard in scientific psychological research will not only help to improve statistical accuracy (and dispel confusion or uncertainty caused by unexplained inconsistencies in research results) but will also help to provide important new insights that can be used to build a richer theoretical understanding of a given research area. This idea relates to Bryan et al.'s (2021) so-called 'heterogeneity revolution', which proposes the use of heterogeneity as a tool for building more complete theories and producing more robust and predictable effects across contexts and populations (rather than simply viewing heterogeneity as a limitation or nuisance associated with the research area). However, replication studies are not always straightforward to conduct as some studies are not reported in sufficient detail to allow for an accurate replication (see William, 2022). In this context, it is worth referring to the findings of the quality assessment presented at the beginning of this review, which found only two of the 71 studies to have been preregistered. Presenting detailed outlines of experimental methods and procedures, in addition to clear outcome measures, prior to publication will allow subsequent researchers to conduct accurate replications of the work. In this vein, preregistration repositories such as the Open Science Framework (OSF) actively encourage those conducting replication studies to contact the authors listed on the initial preregistration to ensure the accuracy of study procedures.

CHAPTER 2 GENERAL DISCUSSION

This systematic review provides the first evidence synthesis of the associations between paranormal beliefs and cognitive function since the early '90s (Irwin, 1993) and the first assessment of study quality. The review identified 71 studies involving 20,993 participants. While most studies achieve good-strong quality ratings, specific areas of methodological weakness warrant further attention. In particular, studies often employ large numbers of measures, metrics and analyses, with no clearly identified primary outcome or adjustment of probability levels. These factors necessarily constrain any firm conclusions because of the high probability of Type 1 errors. Second, information about nonrespondents was either unreported or reported with insufficient detail to permit an assessment of potential nonresponse bias. Finally, up to a third of studies failed to discuss study limitations. The cognitive deficits hypothesis is apparent in most papers (55/71), and a simple vote count shows that two-in-three studies (46/71) document that paranormal beliefs are associated with poorer cognitive performance. The most consistent findings across the six cognitive domains emerged between paranormal belief and an intuitive thinking style, with all eight studies confirming a positive association. Consistent findings also emerged for a bias towards confirmatory and disconfirmatory outcomes, as well as for poorer conditional reasoning ability and perception of randomness, though fewer studies were conducted in these areas. The two studies assessing executive functioning identified a negative association with paranormal belief but showed some inconsistency depending upon the type of executive test used. Associations with all other aspects of cognitive functioning (perceptual decision-making, jumping to conclusions and repetition avoidance, the conjunction fallacy, probabilistic reasoning, critical thinking ability, intelligence, analytical thinking style, and memory) have proven inconsistent, with nearly equal numbers of significant and null findings. Various measurement issues, however, need to be considered. One concerns the large number of paranormal belief measures employed and their varied psychometric properties. The studies reviewed employed 26 different tests of paranormal belief, with the most common being the RPBS and a Rasch variant, with the next most common being 13 bespoke tests created by the authors. Such variability most likely contributes to heterogeneity across studies and potentially undermines the reliability of reported associations between cognitive functions and paranormal beliefs. For a full summary of the scales used in each study, see Appendix I. Not only does the range of cognitive measures used within each cognitive domain contribute to heterogeneity across studies, but so does the reliability of such measures. As Hedge et al. (2018) note, individual differences in relation to cognition and brain function often employ cognitive tasks that have been well-established in experimental research. Such tasks may not be directly adaptable to correlational research, however, for the very reason that they elicit robust experimental effects; they are specifically designed and selected for low between-participant variability. Most studies presented here are correlational and use a combination of established experimental tasks (e.g., the WCST, Raven's Matrices, Cognitive Reflection Test, Embedded Figures Test) and questionnaire-based methods to assess cognition. This may undermine the reliability of reported associations between cognitive functions and paranormal beliefs if studies use experimentally derived cognitive tasks that are sub-optimal for correlational studies. Hedge et al. (2018) offer several suggestions to overcome this, such as the use of alternative statistical techniques (e.g., structural equation modelling), factoring reliability into a-priori power calculations to reduce the risk of bias towards a null effect, or using within-subjects designs when the primary goal of the study is to examine associations between measures rather than focusing on individual differences per se. The largely correlational approach of studies reviewed here also suffers from the standard limitations of questionnaire studies and correlational designs. Although regression approaches can be powerful, they cannot establish causality without the use of longitudinal methods. This correlational approach also means that moderators and mediators of the relationship between paranormal beliefs and cognition remain underspecified.

Future directions – the fluid-executive model

The general trend of the current review accords with the cognitive deficits hypothesis approach described by Irwin (1993) almost 30 years ago—at least insofar as around 60% of published studies document paranormal beliefs to be associated with poorer cognitive performance. Nonetheless, the cognitive deficits hypothesis does not provide an entirely satisfying account of why paranormal believers and sceptics perform differently on such a wide variety of cognitive tasks. This has some key implications: first, that people who believe in the paranormal seemingly have a disparate array of cognitive deficits—are these assumed to have occurred independently of each other or do they somehow accumulate various cognitive deficits? Another implication is that such an array of cognitive deficits is largely atheroetical, with various researchers pursuing seemingly independent lines of research linking cognitive function to paranormal beliefs with little attention to integration. Hence a somewhat underspecified model pervades the literature, with often limited justification for the specific role played by cognitive function in paranormal beliefs or how and why such an array of deficits are identifiable in paranormal believers. Given the almost complete lack of preregistration, accompanied by the large numbers of statistical analyses often conducted without correction, we also cannot exclude concerns about potential publication bias, false positives, and selection bias. Empirical studies presenting significant or favourable findings are, of course, more likely to be published (Song et al., 2009); and crucially, psychologists tend to rate studies as having better quality when they conform to prior expectations. Hergovich et al. (2010) demonstrated this bias by presenting psychologists (all of whom did not believe in astrology) with descriptions of parapsychological studies, finding that they gave higher quality ratings to studies disproving astrological hypotheses. Participants were less likely to complete the study if they received an abstract confirming astrological hypotheses, with an attrition rate of 38.90%. These issues underscore the importance of pre-registered replications of key findings (see Laws, 2016, for a discussion). To our knowledge, potential publication bias has not been extensively assessed. A previous meta-analysis of psychokinesis studies indicated the presence of publication bias (Bösch et al., 2006), but this claim has been challenged (Radin et al., 2006). Finally, questions also arise about whether poorer performance by believers on any cognitive ability tests even merits the descriptor of 'deficits'; and recently has been rephrased more neutrally as the cognitive differences hypothesis (Gray & Gallo, 2016). The term 'deficit' typically implies a permanent lack or loss of cognitive function; however, little to no research has looked at the consistency of cognitive performance in paranormal believers across time and established whether poorer cognitive performance is more trait than state dependent. While paranormal beliefs appear to be largely trait-like, they may have a state component (Irwin et al., 2018).

While current studies do not necessarily endorse Irwin's (1993) comment that ". . .the believer in the paranormal is held variously to be illogical, irrational, credulous, uncritical, and foolish" (p.16), they converge on an underlying non-specific cognitive deficit or collection of deficits. Typically, when an array of cognitive deficits/differences are documented, researchers would want to know if specific areas of cognitive weakness emerge. Currently, no cognitive area suggests a specific deficit profile in paranormal believers.

Although not directly tested, paranormal believers might display heterogeneous cognitive profiles that link to different paranormal belief components. Nonetheless, it is hard to see why or how specific types of paranormal belief content would link to different cognitive deficits. One possibility is that the failure of any specific area of cognitive dysfunction to emerge (amongst perceptual and cognitive biases, reasoning, intelligence, critical thinking and academic performance, thinking style, and executive functioning), may point to a common shared underlying cognitive component. One feasible interpretation is that many of the tasks described in the various domains described here do in fact share a common cognitive ability— higher-order executive functions (planning, reasoning and problem-solving, impulse control, initiation, abstract reasoning, and mental flexibility), which in turn may be related to aspects of fluid intelligence (Diamond, 2013). Human functional brain imagining identifies strikingly similar patterns of prefrontal cortex activity in response to cognitive challenges across various seemingly different domains, including: increased perceptual difficulty (high vs low noise degradation), novelty, response conflict, working memory, episodic and semantic memory, problem solving, and task novelty (Duncan, 2006; Nyberg et al., 2003; Duncan & Owen, 2000). This demand-general activity underlies our ability to engage in flexible thought and problem-solving (Duncan, 2006) and is closely linked to fluid intelligence (Duncan, 2010). It is therefore proposed that the broad cognitive-deficit profile linked to paranormal beliefs may overlap with functions of the multiple-demand (MD) system. Part of the function of the MD system concerns its role in the separation and assembly of task components and that this accounts for the link with fluid intelligence. In this context, it is suggested that each of the cognitive domains linked to paranormal beliefs may indeed be subserved by this MD system housed in the fronto-parietal cortex. The section on executive function is self-evidently linked with the frontal system. The section on intelligence similarly highlights links between paranormal beliefs and fluid IQ measures such as the Ravens Matrices (Raven et al., 2000; Raven, 1976). Studies further show the same MD system is recruited when confronted with perceptually difficult tasks (such as those outlined in the section on perceptual and cognitive biases for degraded visual input, see van Elk, 2015; Simmonds-Moore, 2014; Riekki et al., 2013; Blackmore & Moore, 1994). Aside from supporting our problem-solving ability, fluid intelligence and various aspects of executive functioning (e.g., working memory) underpins our ability to reason and to see relations among items and includes both inductive and deductive logical reasoning. The section on reasoning shows paranormal beliefs are related to conditional and probabilistic reasoning (Denovan et al., 2018; Prike et al., 2017; Dagnall et al., 2016a; Dagnall et al., 2016b; Rogers et al., 2016; Dagnall et al., 2014; Rogers et al., 2009; Dagnall et al., 2007; Bressan, 2002; Musch & Ehrenberg, 2002; Robers & Seager, 1999; Blackmore, 1997; Brugger et al., 1990; Wierzbicki, 1985). Thus, many of the cognitive deficit-paranormal belief associations may be reframed as the product of a single underlying fluid intelligence-executive component. Going forward, such a model suggests potential avenues of research. One prediction would be that groups of believers and sceptics matched for fluid IQ would be less likely differ on a range of cognitive tasks.

Limitations of the present review

The current review is the first to assess the quality of studies examining cognitive function and paranormal beliefs. Study quality is good-to-strong, with interrater reliability on AXIS ratings being almost-perfect (93%). Individual AXIS items however are not weighted and any simple comparisons between specific studies across total summed quality scores should be regarded with caution (Greenland & O'rourke, 2001; Jüni et al., 1999; Greenland & Robins, 1994). Thus, two studies with the same total quality score, but across different items, might not be comparable because some items may be more concerning to quality than others. Hence, the review focuses on specific domains of strength or weakness across studies. There are, however, substantial limitations regarding the classification of studies into six areas of cognitive function: (1) perceptual and cognitive biases, (2) reasoning, (3) intelligence, critical thinking, and academic performance, (4) thinking style, (5) executive function, and (6) other cognitive functions. Appendix B shows that many of the studies could be re-classified and indeed, two-thirds (48/71) could be re-classified as assessing executive functioning. The latter is consistent with the proposal that a substantial proportion of the published studies may be documenting a relationship between paranormal beliefs and higher-level executive function/ fluid intelligence. The preregistered protocol had an exclusion criterion concerning samples with individuals aged less than 18, and this led to the exclusion of 11 datasets (see Appendix A for a complete list and details; Aarnio & Lindeman, 2005; Saher & Lindeman, 2005; Lindeman & Aarnio, 2006) were overlapping or identical samples). A key reason for exclusion was because age impacts both cognitive functions and paranormal beliefs. Certain cognitive functions, for example executive functions, take until late adolescence or early adulthood to mature (Ferguson et al., 2021). Additionally, younger individuals also show higher levels of paranormal beliefs (Emmons & Sobal, 1981; for a discussion see Irwin, 1993). While the exclusion of these studies is a potential limitation, their exclusion does not change the key findings or conclusions drawn from this review. In the same context, the lack of an upper age limit exclusion criterion could also be considered as a limitation. Sixteen papers (23%) reviewed here included participants aged 65+ (though 25/71 (36%) studies did not report on the age range of participants). While some cognitive functions do not mature until late adolescence or early adulthood, measurable changes in cognitive function occur with normal aging. Performance on certain cognitive tasks has been shown to decline with age, such as those requiring executive functioning (including decision-making, working memory and inhibitory control), visuoperceptual judgement and fluid-intelligence (Murman, 2015; Salthouse et al., 2003). Such cognitive declines have been associated with age-related reductions of white matter connections in brain regions including the prefrontal cortex (Kennedy & Raz, 2009; Tisserand & Jolles, 2003). Finally, one limitation is that a meta-analysis could not be conducted because of the large variability in outcome measures within and between studies, which make it challenging to determine the precise outcome being tested. In parallel, the large numbers of analyses per study also mean that conclusions from this systematic review regarding findings for specific cognitive domains must also be interpreted with some caution.

Conclusions

The systematic review identified 71 studies spanning: perceptual and cognitive biases, reasoning, intelligence, critical thinking, and academic performance, thinking styles, and executive function. However, then tasks employed to assess performance in each domain often appear to require higher-order executive functions and fluid intelligence. A new, more parsimonious, fluid-executive theory account is therefore proposed for future research to consider. Methodological quality is generally good; however, specific theoretical and methodological weaknesses have been identified within the research area. In particular, it is recommended that future studies preregister their study design and proposed analyses prior to data collection, and address both the heterogeneity issues linked to paranormal belief measures and the reliability of cognitive tasks. It is hoped that these methodological recommendations alongside the fluid-executive theory will help to further progress current understanding of the relationship between paranormal beliefs and cognitive function.

CHAPTER 3. MEASURING PARANORMAL BELIEFS

INTRODUCTION

The Revised Paranormal Belief Scale

The Paranormal Belief scale in both original (PBS; Tobacyk & Milford, 1983) and revised format (RPBS; Tobacyk, 2004) is the most widely used measure of paranormal beliefs. The original PBS contained 25-items measuring belief across seven domains: Traditional Religious Beliefs (TRB), Psi, Witchcraft, Superstition, Spiritualism, Extraordinary Life Forms (ELF), and Precognition. Subjects rated their level of agreement with items such as “there are actual cases of Voodoo death” using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). While the PBS demonstrated a high test-retest reliability over a four-week period (.89, Tobacyk & Milford, 1983), Tobacyk (2004) later revised the scale to address some of the shortcomings associated with the response format and item content of the original PBS. The RPBS extended the original 5-point scale to a 7-point scale to provide a less restrictive measure of subjects’ agreement with each item. The most substantial change seen in the RPBS, however, is the removal of six items across three subscales, and the addition of a further seven items. This resulted in the 26-item RPBS (for a more detailed summary of the item changes made in the development of the revised scale, see Table 3.1.). The six items that were replaced in the RPBS were removed for either ambiguous phrasing (e.g., “some people have the ability to accurately predict the future” could be considered to apply to those with specific scientific training, such as a meteorologist accurately predicting tomorrow’s weather), or were removed to improve cross-cultural validity (e.g., subjects from some countries may be unfamiliar with the concept of “Bigfoot”, as familiarity with this cryptozoological creature is arguably largely restricted to those from America). The remaining items from the PBS were retained in the RPBS, as were the original seven factors. The changes made to the RPBS raised the four-week test-retest reliability of the scale to .92 (Tobacyk, 2004).

Table 3.1. Summary of the six items removed from the PBS and their seven replacements in the RPBS.

Subscale	Removed PBS Item	New RPBS Item
Precognition	1 Dreams can provide information about the future. 2 Some people have the ability to predict the future. 3 The idea of predicting the future is foolish.	7 Astrology is a way to accurately predict the future. 14 The horoscope accurately tells a person's future. 21 Some psychics can accurately predict the future. 26 Some people have an unexplained ability to predict the future.
Witchcraft	10 Voodoo is a real method to use paranormal powers. 24 There are actual cases of Voodoo death.	17 Through the use of formulas an incantations, it is possible to cast spells on persons. 24 There are actual cases of witchcraft.
ELF	20 Bigfoot exists.	20 There is life on other planets.

Note: Original item number for the PBS items, and the item numbers for the RPBS replacements, are provided next to each item in the table

While the RPBS presents several improvements over the original PBS and has seen widescale use, several issues have been raised regarding the factor structure and item content of the revised scale. Despite the changes made to improve cross-cultural validity of the items, the scale still proves problematic in samples with different cultural backgrounds. Many of these cultural criticisms are focused on the cryptozoological content of the ELF subscale. When assessing the validation of the RPBS in French samples, Bouvet et al. (2014) found the ELF subscale, with items relating to the Loch Ness monster and the abominable snowman of Tibet, to have a low internal consistency suggesting that this dimension of belief is not particularly relevant in French samples. In attempts to address the cultural shortcomings of the ELF subscale items, several studies report replacing these items with equivalents more relevant for their samples. For example, in his research with Turkish university students, Dag (1999) replaced “abominable snowman of Tibet” with “wolfman” and “Loch Ness monster of Scotland” with “Van Lake monster of Turkey”. Similarly, when using the original PBS with a South African student sample, Peltzer (2003) replaced “abominable snowman of Tibet” with “Tokoloshi”, “Loch Ness monster of Scotland” with “Sahuwe”, and “Big Foot” with “zombies”. Some studies using American samples have also removed the culture-bound items of the ELF subscale when using the RPBS, due to a lack of familiarity with these cryptozoological creatures (see Willard & Norenzayan, 2013).

The TRB subscale has also raised concerns due to contradictory evidence concerning the relationship between paranormal and religious beliefs, leading many researchers to question whether religious items should be included in paranormal scales. While several studies have noted positive correlations between the two belief types (see Lindeman & Svedholm- Häkkinen, 2016; Hergovich et al., 2005; Orenstein, 2002), others have found those displaying especially strong forms of religious belief to be less likely to endorse the existence of paranormal phenomena (see Beck & Miller, 2001; Hillstrom & Strachan, 2000). Baker and Draper (2010) and Bader et al. (2012) provide a potential explanation for these conflicting findings, suggesting that the relationship between religious and paranormal beliefs may be better conceptualised as curvilinear. In this model, paranormal beliefs increase alongside religious beliefs, but then *decrease* when religious beliefs become particularly strong.

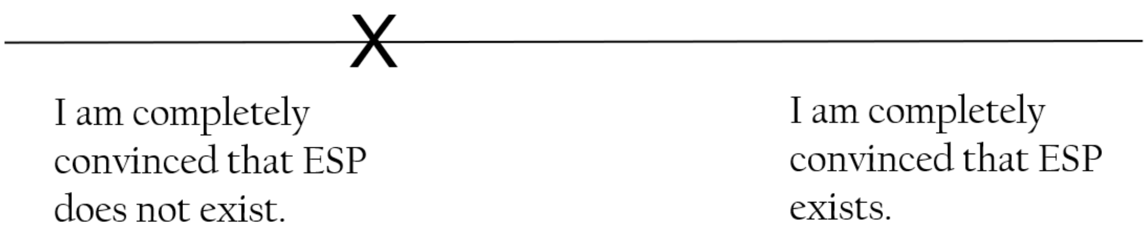
A further criticism of the item content within both the PBS and RPBS surrounds the lack of negatively worded items, with only two featured in the original PBS (“mind reading is not possible” and “the idea of predicting the future is foolish”) and one in the RPBS (“mind reading is not possible”). The lack of negatively worded items on both scales increases the risk of scores being affected by subjects endorsing the items without fully considering the content (see Furr, 2011, pp. 16-24 for a discussion on negatively worded questionnaire items).

Finally, criticisms of both the PBS and RPBS have focused on the factor structure of the scales, which presents seven dimensions of paranormal belief. The question of how many paranormal belief dimensions exist has been the centre of much debate, and perhaps stems from the lack of an agreed-upon definition for paranormal phenomena. Many studies have attempted to replicate the large seven-factor structure of both the PBS and RPBS. While some (e.g., Drinkwater et al., 2017; Haraldsson & Houtkooper, 1996) have replicated this structure, others have failed to do so and instead report smaller factor structures ranging from one to six (see French & Stone, 2013, pp. 13-14).

The Australian Sheep-Goat Scale

Another widely used measure of paranormal beliefs is the Australian Sheep-Goat Scale (ASGS; Thalbourne & Delin, 1993). The ASGS is a shorter 18-item measure with three subscales: Belief in Extrasensory Perception (ESP), Psychokinesis, and Life After Death. In the original ASGS, subjects rated their level of agreement with each item using a visual analogue scale. Therefore, each item is presented to subjects with two types of phrasing, one which represents belief in the phenomena described and one which represents disbelief (i.e., scepticism). See Figure 3.1. below for an example of the visual analogue scale, including item phrasing and an example of a possible subject response.

Figure 3.1. Example of the ASGS visual analogue response scale.



Note: ASGS = Australian Sheep-Goat Scale, ESP = extrasensory perception.

The visual analogue scale used in the original ASGS uses a ruler to convert subjects' marks to scores ranging from 1-44, with higher scores indicative of higher paranormal belief. These scores are then transformed (1-10 = 0, 11-30 = 1, and 31-44 = 2), and are used to produce a total score with a possible maximum of 36. However, this response scale raises concerns surrounding the reliability of subjects' scores. As Roe (1998) argues, some subjects may find it unclear when expressing complete (dis)belief as to whether they should place an "x" at the line pole, at the centre of the text, or at the left edge of the text. To address the issues of the visual analogue scale, subsequent studies have since employed different response methods when using the ASGS. Roe (1998) addressed his own concerns by replacing the visual analogue scale with a six-point Likert scale (1 = strongly disagree, 6 = strongly agree). Thalbourne's (1995) forced-choice format also offers a simpler response format compared to the original visual analogue ("true", "uncertain", and "false", coded with 2, 1 or 0 points respectively), but results in a similar total ASGS score to the original coding. However, a strength of the original visual analogue format is that it presents both negatively and positively worded items, a feature which is lacking from the subsequent adaptations of the ASGS. Therefore, much like with the PBS and RPBS, these adapted versions of the ASGS raise concerns about response biases. Despite their differences, each version of the ASGS has demonstrated high internal reliability, with Cronbach's alpha values between .82 and .96 (see Prike et al., 2018; Stone, 2016; Brotherton & French, 2014; Thalbourne, 2010; Voracek, 2009). The scale has also shown positive correlations of 0.70 and above with the RPBS (see Storm et al., 2017; Dagnall et al., 2014).

The Survey of Scientifically Unaccepted Beliefs

The Survey of Scientifically Unaccepted Beliefs (SSUB, also referred to as the "Survey of Popular Beliefs"; Irwin & Marks, 2013) is a more recent scale used to assess belief in paranormal phenomena. The SSUB was reduced from an initial collection of 92 items to a 20-item self-report measure containing two subscales: New Age Beliefs and Traditional

Religious Beliefs. Although not as widely used (or scrutinised) as the RPBS or ASGS, the SSUB has shown high levels of internal reliability with Cronbach's alphas at and above .89 for the subscales and total scale, both in the original study (Irwin & Marks, 2013) and in subsequent studies (see Irwin et al., 2018; Irwin et al., 2015). The scale also has a more balanced composition of positive and negative items compared to the RPBS or ASGS. While many of the phenomena contained within the SSUB could be considered paranormal (e.g., the existence of genuine haunted houses, the accuracy of horoscopes, and the existence of various psychic abilities), the scale also contains items relating to several scientifically unaccepted beliefs that are not commonly associated with the paranormal such as crop circles and pixies. Such phenomena are based upon mystery and elusiveness rather than a strict violation of scientific principles, with this violation often used to define phenomena as "paranormal" (see Lawrence, 1995).

Other scales

While the PBS, RPBS, ASGS and SSUB have been identified as the most widely used measures of paranormal beliefs, they do not represent an exhaustive list of all existing paranormal belief scales. For example, the Manchester Metropolitan University New (MMU-N, unpublished; see Dagnall et al., 2014; Dagnall et al., 2010a, 2010b) is a 47-item scale tapping into eight factors of paranormal belief: Hauntings, Superstitions, Religious Belief, Alien Visitation, Extrasensory Perception, Psychokinesis, Astrology, and Witchcraft. Subjects rate their level of agreement with items (e.g., "poltergeists exist") using a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). All factors demonstrate high internal reliabilities, with all Cronbach's alpha's above .80 (Dagnall et al., 2010a). Žeželj et al. (2009) developed a shorter Superstitious Beliefs and Behaviour Scale (SSBS) consisting of 20 items relating to Good Consequences, Bad Consequences and Fortune Telling. Subjects rate their level of agreement with SSBS items (e.g., "knocking on wood protects me from bad things happening") using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The scale demonstrates a high internal reliability with a Cronbach's alpha of .87 (Žeželj et al., 2009). Other scales that have seen some use in paranormal research include the Anomalous Experiences Inventory (Gallagher et al., 1994), Belief in the Paranormal Scale (Jones et al., 1977) and the Occultism Scale (Böttinger, 1976). Despite the wide range of available measures, many studies rely on their own novel measures to assess paranormal beliefs within their samples. Griffiths et al. (2019) developed a novel 25-item scale for use in their study, with subjects rating their level of agreement with items such as "what happens to people is determined by fate" on a 5-point Likert scale (0 = strongly disagree, 4 = strongly agree). Similarly, Rizeq et al. (2020) used a novel 35-item scale with items adapted from several existing paranormal belief measures. In their study, Rizeq et al. (2020) asked subjects to rate their level of agreement with items (e.g., "certain types of crystals have special powers") on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree). Bressan (2002) created a 20-item paranormal belief scale, with subjects rating their level of belief in each item (e.g., "amulets keep negative forces away") using a 7-point Likert scale (1 = I do not believe it, 7 = I believe it). Other studies have adopted a shorter range of questions for assessing paranormal beliefs within their samples, ranging from six items to sixteen (see Majima, 2015; Orenstein, 2002; Blackmore & Moore, 1994; Alcock & Otis, 1980). As much of the research into paranormal beliefs employs a vast range of measures, and often includes "one-off" novel scales developed for specific use in a single study, careful attention must be paid when making direct comparisons between results from different studies, and when considering the validity and replicability of such studies.

Differential item functioning

Some researchers have questioned whether variations in responses on paranormal belief scales represent a true fluctuation in belief level, or whether these variations may be partly a function of semantic biases introduced by age or gender. This issue is commonly referred to as differential item functioning (DIF). Some authors have applied Rasch scaling (a modern test theory model) to existing paranormal belief scales as a way of detecting these biases and assessing their effect. Lange and Thalbourne (2002) applied this method to the ASGS and found the scale to be a reliable and unbiased measure of belief in paranormal phenomena in a large sample of nearly 2,000 respondents. Weak age and gender differences were found, but these biases had minimal impact on the data gathered from the ASGS (Lange & Thalbourne, 2002). In a similar study, Lange et al. (2000) applied Rasch scaling to the RPBS and found significant DIF for gender on 18 items, and significant DIF for age on 15 items. Consequently, Lange et al. (2000) used a "top-down" purification method to address the biases seen in the RPBS. Specifically, the authors first conducted principal component analysis before "purifying" the scale by developing subscales free of DIF. This "top-down" method (combining factor analysis, Rasch scaling, and statistical tests of DIF and item dimensionality) resulted in a two-factor model for the RPBS which presented reduced DIF (Lange et al., 2000). This two-factor model of the RPBS has been used in several subsequent studies and is considered by some as an improved version of the RPBS (see Drinkwater et al., 2012; Irwin et al., 2012; Watt et al., 2007; Rogers et al., 2006; Terhune & Smith, 2006). Despite the extensive use of the purified scale, several items failed to load on either of the two new factors, with the authors highlighting that addition of new items to the RPBS may produce additional belief clusters to those identified through their analyses (Lange et al., 2000). DIF analysis was used during construction of the SSUB, with three items from the original item pool identified for biases of age and gender. These items were therefore able to be removed from the final scale to limit any effects of DIF on the SSUB (Irwin & Marks, 2013).

Scale Development - Classical versus modern test theory

Latent traits such as paranormal beliefs are, by definition, unobservable. Therefore, research relies on the use of self-report scales, like those detailed above, which assume that individuals' responses to items are influenced by the latent trait of interest (Sharkness & DeAngelo, 2011). Classical test theory (CTT) and modern test theory (MTT; also referred to as item response theory) are the two primary methods used in psychological scale development. Both CTT and MTT models strive to measure and improve the reliability, validity, and internal consistency of the scale under assessment (Rusch et al., 2017; Magno, 2009) but do so in different ways. One of the key differences between these approaches is that CTT assumes that measurement precision is equal for all individuals, while MTT takes the view that measurement precision depends on individuals' levels of the latent trait (Jabrayilov et al., 2016).

CTT models, focused at the test-score level, assume a linear model that links the observable test score (X) to the sum of two unobservable variables: true score (T) and error score (E) (Hambleton & Jones, 1993). This assumption can be more clearly illustrated with the following formula: $X = T + E$. In this formula, the observed score (X) represents the observed total score calculated from the scale in use, and the error score represents a random, non-systematic error assumed to be independent of the true score (e.g., poorly functioning test items, or external confounding variables). The true score is often conceptualised as the mean of all scores obtained if an individual responded to the given scale an infinite number of times (Downing, 2003). Therefore, the observed score of X can be considered to be a combination of both relevant information relating to the latent variable of interest and the error associated with each item (DeVellis, 2006). A factor-analytic strategy (often relying on the use of exploratory factor analysis for item selection) is among the most popular CTT method for scale development, and has the primary aim of developing an internally consistent scale with a manageable number of differentiable dimensions (Simms, 2008).

CTT models offer certain advantages. For example, many CTT models are based on relatively weak assumptions, and are therefore easily met with real test data (Hambleton & Jones, 1993). These models are also simple to use and allow for examination (at the test-score level) of the precision with which the latent trait of interest is measured by a given scale (De Champlain, 2010). However, CTT's standing popularity, despite the emergence of more modern approaches to scale development, could be attributed to the fact that many researchers are familiar with its basic concepts and are likely to have encountered CTT (or to have used scales that were developed through CTT methods; DeVellis, 2006). Therefore, it is important to also consider the limitations of CTT. The central limitation of CTT models is that person and item parameters are sample-dependent, which limits the utility of these statistics in scale development (Magno, 2009; Hambleton & Jones, 1993). CTT models also do not allow for rigorous assessment of item characteristics that can be computed under different models, and so scales developed using CTT methods may suffer from differential item functioning (DeVellis, 2006).

In contrast to CTT models, MTT models are nonlinear and focus at the item level, seeking to relate respondents' performance on individual test items to their estimated level of the latent trait of interest (Urbina, 2014). These models are assumed to be invariant across populations, meaning the item and test parameters can be interpreted independent of specific samples. The type of MTT model used in scale development may differ depending on the type of data collected (dichotomous data such as yes/no responses, or polytomous data collected using Likert response methods), and on the number of dimensions they specify. In general, MTT models can be said to have three main goals: (1) to produce items that provide the most information about respondents' levels of the latent trait of interest, (2) to present respondents with items tailored to their latent trait levels, and (3) to reduce the number of items needed to determine respondents' level of the latent trait without loss of reliability (Urbina, 2014). The advantages of MTT models over CTT models are most notable at the item level. Item characteristics, differential functioning and fit to the model can be assessed, as well as individuals' response styles and the functionality of response scales (Kline, 2005). However, a limitation of MTT models is their use of sophisticated and in-depth statistical analyses which remain unfamiliar to many researchers and testing professionals (Urbina, 2014). The assumptions of MTT models are also more restrictive compared to those of CTT models (i.e., more difficult to meet with real test data), and sample size requirements are much larger for both items and respondents (Kline, 2005). For unidimensional MTT models (such as the Rasch model), minimum sample sizes of approximately 200 respondents are required (Downing, 2003). However, multidimensional MTT models require large sample sizes ≥ 1000 respondents to identify precise item parameters and decrease error estimation (Kose & Demirtasli, 2012).

The present studies

The present studies build upon the vast literature surrounding the psychometric assessment of paranormal beliefs within the general population, with a particular focus on the shortcomings of the RPBS and the improvement of paranormal belief measurement.

As several studies have reported difficulties in replicating the original factor structure of the RPBS, the first study had two main aims: to replicate the original seven factors through confirmatory factor analysis, and to determine whether a smaller, more appropriate factor structure could be devised through exploratory factor analysis. While there were no specific predictions regarding the number of factors for the exploratory factor analysis, it was anticipated that the results of the confirmatory factor analysis would not support the same factor structure as that presented by Tobacyk (2004), owing to the replication difficulty seen in previous studies.

The second study aims to develop a new measure of paranormal beliefs that addresses some of the shortcomings observed in existing scales, and therefore provides a more reliable measure of paranormal beliefs within the general population. The study sought to address these issues through the inclusion of both positively and negatively phrased items, and items solely concerned with phenomena widely considered to be "paranormal". The study also aimed to include paranormal phenomena missing from current scales, such as hauntings and good luck. In addition, the study aimed to ensure the scale was less culture-bound than existing scales, did not include many religious items, and did not contain evidence of significant differential item functioning. The study was conducted in two parts, with the construction of the new scale and examination of its latent structure assessed using both classical (factor analysis) and modern (Rasch analysis) test theories.

STUDY 3.1. CONFIRMATORY AND EXPLORATORY FACTOR ANALYSIS OF THE REVISED PARANORMAL BELIEF SCALE

METHOD

Participants

An opportunistic sample of the general public (*N*=238) was recruited through word of mouth and advertisements placed on social media. Inclusion criteria were being fluent in English and aged 18 or over. During data cleaning, 71 incomplete responses (29.83%) were removed from the dataset. The final sample (*N*=167) consisted of 100 females (59.90%) and 67 males (40.10%), with most participants aged between 25-34 (27.50%). Most participants were white (68.50% white British, 16.80% other white background, 04.80% white Irish) and were educated to postgraduate level (27.50%). Of the participants with a university education (27.50% postgraduate degree, 19.20% undergraduate degree, 13.20% doctoral degree), most had a background in psychology (18.00%). See Table 1.2. for full demographic details.

Materials

The only material included in the analyses of this study was Tobacyk's (2004) Revised Paranormal Belief Scale (RPBS; see Appendix J). The scale consists of 26 items (e.g., "there is a devil") and seven subscales. Subjects rate their level of agreement with each item using a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). Ratings for item 23 ("mind reading is not possible") require reverse coding prior to analyses. A total RPBS score is calculated by totalling the ratings across all items. Scores can also be calculated for each of the seven subscales using the mean of the subscale items: Traditional Religious Belief (TRB; mean of items 1, 8, 15 and 22), Psi (2, 9, 16, 23), Witchcraft (3, 10, 17, 24), Superstition (4, 11, 18), Spiritualism (5, 12, 19, 25), Extraordinary Life Forms (ELF; 6, 13, 20), and Precognition (7, 14, 21, 26). Tobacyk (2004) reports a mean RPBS score of 89.10 (*SD* = 21.90) in a sample of 217 respondents. High Cronbach alpha reliability statistics have been identified for the total scale, with an alpha of .93 reported by Drinkwater et al. (2017), and .94 reported by Wilson et al. (2014).

Procedure

The questionnaire and demographic items were administered as part of an online survey using Qualtrics Survey Software (Qualtrics, Provo, UT; https://www.qualtrics.com). Respondents who agreed to take part were asked to provide their age, gender (male, female, other), ethnicity (Arabic, Asian/Asian British, Bangladeshi, Black/Black British, Chinese, Indian, Pakistani, White British, White Irish, other Asian background, other White background, mixed background) level of education (doctoral degree, postgraduate degree, undergraduate degree, post-secondary education, secondary education, vocational) and academic discipline if they had indicated a university education (architecture, arts and humanities, business, education, law, medicine, natural sciences, philosophy, psychology, social sciences, theology, technology, other medical, other). Respondents were not required to provide any of the above demographic details. The RPBS was then presented, with respondents' level of agreement with each item recorded using a 7-point Likert scale (Strongly Disagree, Moderately Disagree, Slightly Disagree, Uncertain, Slightly Agree, Moderately Agree, Strongly Agree). The seven response options were numerically coded from 1 to 7 for the 25 positively worded items, and reverse coded for the one negatively worded item.

Informed consent was obtained from all participants and ethical approval for the study was granted by the University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (HSET ECDA; protocol number LMS/PGR/UH/03844).

RESULTS

Table 3.2. presents a summary of participant demographic information. Of the final sample, most were white, well-educated females.

Table 3.2. Frequencies and percentages of participant demographics.

Variable	n	%
Gender		
Male	67	40.10
Female	100	59.90
Age		
18-24	17	10.2
25-34	46	27.5
35-44	34	20.4
45-54	38	22.8
55-64	27	16.2
Over 65	5	3.0
Ethnicity		
Prefer not to say	1	0.6
Asian/Asian British	1	0.6
Black/Black British	1	0.6
Indian	2	1.2
Pakistani	1	0.6
White British	115	68.9
White Irish	8	4.8
Mixed background	7	4.2
Other White background	28	16.8
Other Asian background	3	1.8
Education		
Prefer not to say	2	1.2
No formal education	1	0.6
Primary education	1	0.6
Vocational	10	6.0
Secondary education	10	6.0
Post-secondary education	43	25.7
Undergraduate degree	32	19.2
Postgraduate degree	46	27.5
Doctoral degree	22	13.2
Discipline		
Art & humanities	8	4.8
Business	4	2.4
Education	5	3.0
Law	2	1.2
Medicine	14	8.4
Natural sciences	8	4.8
Other medical	8	4.8
Philosophy	3	1.8
Psychology	30	18.0
Social sciences	8	4.8
Technology	7	4.2
Other	14	8.4
N/A (have not attended university)	56	33.5

Descriptive statistics

Means, standard deviations, minimum and maximum scores are presented in Table 3.3. for both the present sample and Tobacyk's (2004) original sample. As shown in Table 3.3., the data suggests an overall lower mean RPBS score for the present sample, and a higher standard deviation compared to the original sample in Tobacyk's (2004) study.

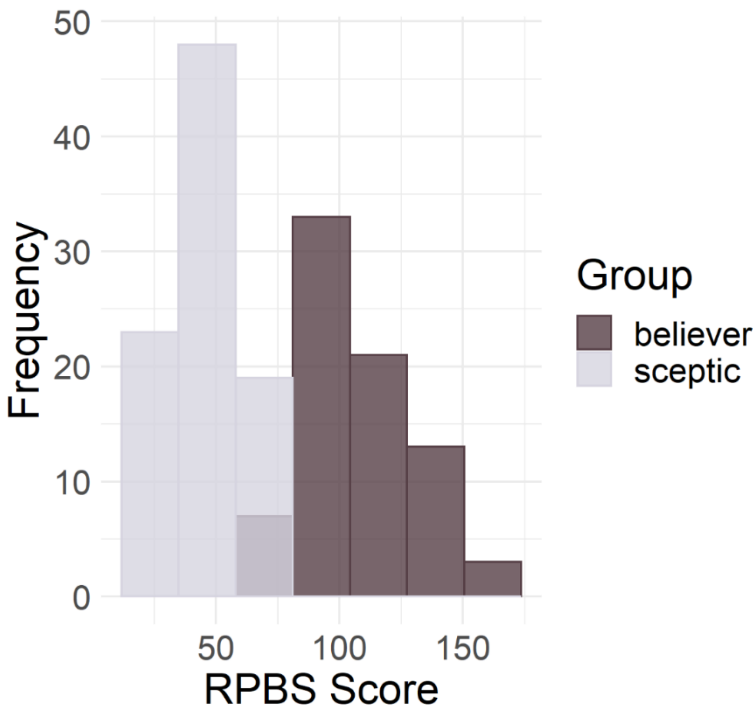
Table 3.3. A comparison of the descriptive statistics on the RPBS for the present sample and Tobacyk's (2004) sample.

Score	Tobacyk's (2004) Sample (*N*=217)				Present Sample (*N*=167)			
	M	*SD*	Min	Max	*M*	*SD*	Min	Max
Full Scale	89.10	21.90	40	165	74.80	36.38	27	166
Subscales								
TRB	6.30	1.20	1	7	3.25	1.92	1	7
Psi	3.10	1.50	1	7	2.87	1.56	1	7
Witchcraft	3.40	1.70	1	7	3.07	1.93	1	7
Superstition	1.60	1.20	1	5	1.66	1.15	1	6.67
Spiritualism	2.80	1.40	1	7	3.16	2.06	1	7
ELF	3.30	1.30	1	7	3.34	1.25	1	7
Precognition	3.00	1.30	1	7	2.60	1.65	1	7

Note: RPBS = Revised Paranormal Belief Scale, TRB = traditional religious beliefs, ELF = extraordinary lifeforms.

These findings suggest that, while overall paranormal beliefs were lower in the present sample, there is more variation in the level of belief for the present sample than the original sample. The minimum score for the present sample was also much lower than in the original Tobacyk (2004) sample, suggesting the present sample were more sceptical of paranormal phenomena. To assess this further, participants were separated into "believers" and "sceptics", as defined by their mean overall RPBS score. Therefore, participants with a mean RPBS score > 74.80 were identified as believers, and those with scores < 74.80 were identified as sceptics. This produced a sample of 77 believers and 90 sceptics, and supports the trend seen in Table 3.3. which suggests the present sample was more sceptical of paranormal phenomena. A visual representation of the distribution of RPBS scores between believers and sceptics in the present sample can be found in Figure 3.2.

Figure3.2. Distribution of total RPBS scores for believers and sceptics.



Note: RPBS = Revised Paranormal Belief Scale

Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) was conducted to determine whether the original seven-factor orthogonal model of the RPBS presented by Tobacyk (2004) could be replicated with the data from the present sample. The model specified the same original seven factors with their original item structures. Model fit was calculated using the lavaan (Rosseel, 2012) package in R version 4.0.2 (R Core Team, 2020). To determine the strength of the model fit, four main fit indices were used: comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardised root mean square residual (SRMR). For both the CFI and TLI, a value of .95 and above would indicate very good model fit, while a value of .90 or above would indicate acceptable fit. For the RMSEA a value of .05 or below would indicate close fit, with .08 or above suggesting acceptable fit. If the *p*-value for the RMSEA is greater than the standardised .05 alpha value, this would confirm a close model fit as indicated by an RMSEA value of .05 or below. Finally, an SRMR value of .05 or below would also indicate a well-fitting model.

The model demonstrated poor fit, with a CFI of .88 and a TLI of .86. RMSEA also indicated a poor model fit with a value of .099 90%CI(.091, .108), as did the SRMR with a value of .07. Despite the poor model fit, all items demonstrated significant factor loadings (all *p*-values <.001). There were also significant covariances between all seven factors. As the original factor structure of the PBS and adopted by the RPBS was described as orthogonal by the authors (see Tobacyk, 2004; Tobacyk & Milford, 1983), a second CFA was conducted to compare the first model to a model treating each factor as independent. Therefore, the second model was calculated without covariances, but using the same seven factor structure. The same four fit indices were used to assess the strength of fit for the second model. Model two demonstrated poor fit, as indicated by a CFI value of .66, a TLI value of .63, and an RMSEA value of .164 90%CI (.156, .171). However, the SRMR value indicated model two was a well-fitting model for the data, with a value of .44. Again, all items loaded significantly onto their designated factors, with all *p*-values ≤.001. For a comparison of the fit indices for both models, see Table 3.4. Finally, a direct comparison of the two models suggested that model one (non-orthogonal, covariances allowed) fit the data significantly better than model two (orthogonal, covariances not allowed), $\chi^2(21)=899.65$, $p < .001$.

Table 3.4. Summary of fit indices for Model 1 and Model 2.

Model Fit Indices	Model 1 (non-orthogonal)	Model 2 (orthogonal)
CFI	.88	.66
TLI	.86	.63
RMSEA	.099***	.164***
SRMR	.07	.44†

*Note: CFI = Comparative fit index, TLI = Tucker-Lewis index, RMSEA = Root mean square error of approximation, SRMR = Standardised root mean square residual. *** = $p < .001$, † = statistic indicating good model fit.*

Exploratory Factor Analysis

Following the CFA, an exploratory factor analysis (EFA) was conducted using SPSS Version 26 to further investigate the latent constructs underpinning the RPBS and to determine whether a smaller factor structure may describe the scale more appropriately. A principal components extraction method was employed and only eigenvalues greater than one were extracted. As the CFA indicated that a non-orthogonal model fit the data significantly better than an orthogonal model, a direct oblimin rotation was used in the EFA. A Kaiser-Meyer-Olkin value of .93 and a significant Bartlett's Test of Sphericity ($\chi^2 = 3989.08$, $p < .001$) indicated that the data were suitable for further analysis. A five-factor solution was extracted and accounted for 74.17% of the total variance. Cronbach's Alpha was computed for each factor, which indicated a poor internal consistency for the fifth factor ($\alpha = .39$). Therefore, the two items comprising the fifth factor (#6: "the abominable snowman of Tibet exists" and #20: "there is life on other planets") were removed and the analysis was conducted again. The second analysis, conducted on 24 items, indicated a four-factor solution and accounted for 73.83% of the total variance. All emergent factors demonstrated acceptable levels of internal consistency and were conceptually distinct (see Table 3.5.).

Table 3.5. Factors with corresponding Cronbach's Alpha (α) scores and item loadings.

Factor	α	Items (loading scores)
1 Spiritual Beliefs	**.97**	2 Some individuals are able to levitate (lift) objects through mental forces (.92) 3 Black magic really exists (.68) 5 Your mind or soul can leave your body and travel (astral projection) (.87) 7 Astrology is a way to accurately predict the future (.70) 9 Psychokinesis, the movement of objects through psychic powers, does exist (.89) 12 During altered states, such as sleep or trances, the spirit can leave the body (.86) 13 The Loch Ness monster of Scotland exists (.59) 14 The horoscope accurately tells a person's future (.73) 16 A person's thoughts can influence the movement of a physical object (.88) 17 Through the use of formulas and incantations, it is possible to cast spells on persons (.78) 19 Reincarnation does occur (.80) 21 Some psychics can accurately predict the future (.86) 24 There are actual cases of witchcraft (.59) 25 It is possible to communicate with the dead (.81) 26 Some people have an unexplained ability to predict the future (.74)
2 TRB	**.89**	1 The soul continues to exist though the body may die (.53) 8 There is a devil (.78) 15 I believe in God (.96) 22 There is a heaven and a hell (.91)
3 Superstition	**.82**	4 Black cats can bring bad luck (.83) 11 If you break a mirror, you will have bad luck (.79) 18 The number "13" is unlucky (.76)
4 Supernatural Powers	**.52**	10 Witches do exist (.61) 23* Mind reading is not possible (.66)

*Note: TRB = traditional religious beliefs, * = reverse scored item*

The first factor, eigenvalue 13.19, accounted for 54.97% of the variance and demonstrated an excellent internal reliability (α = .97). The 15 items contained within Factor 1 included phenomena such as astral projection, astrology, and black magic. Therefore, Factor 1, containing just over 60% of the total scale items, was labelled "*Spiritual Beliefs*". The second factor, eigenvalue 1.80, accounted for 7.52% of the variance and again demonstrated an excellent internal reliability (α = .89). Factor 2 comprised the original four items that made the TRB subscale in the RPBS, and so this factor retained the label "*Traditional Religious Beliefs*". Factor 3 had an eigenvalue of 1.56 and accounted for 6.50% of the variance. The factor had an excellent internal reliability (α = .82) and comprised the same three items as the Superstition subscale in the RPBS. Therefore, Factor 3 retained the original label "*Superstition*". The final factor, eigenvalue 1.16, accounted for 4.84% of the variance and demonstrated low to moderate internal reliability (α = .52). Factor 4 comprised only two items (#10: "witches do exist" and #23: "mind reading is not possible") and was labelled "*Supernatural Powers*". The total 24-item scale demonstrated an excellent overall internal reliability (α = .96).

DISCUSSION OF STUDY 3.1.

An initial confirmatory factor analysis of the RPBS identified poor fit for the original seven-factor solution proposed by Tobacyk (2004), despite significant factor loadings for all items. An exploratory factor analysis with a non-orthogonal rotation method reduced the original 26-item scale to 24 items, describing paranormal belief on four dimensions: *Spiritual Beliefs*, *Traditional Religious Beliefs* (*TRB*), *Superstition*, and *Supernatural Powers*. The first factor (*Spiritual Beliefs*), containing more than half of the total scale items, demonstrated the highest internal reliability (similar to the internal reliability statistic identified for the total 24-item scale), though the internal reliability statistics for the second (*TRB*) and third (*Superstition*) factors were also high. The final factor (*Supernatural Powers*), containing only two scale items, demonstrated a considerably lower internal reliability. The findings from the exploratory factor analysis suggest that paranormal beliefs might be conceptualised as a hierarchical construct, with an overarching belief related to several distinct elements. This accords with previous research that has used the RPBS as both a general and multi-dimensional measure of paranormal beliefs (see Dagnall et al., 2016; van Elk, 2013; Watt et al., 2007; Aarnio &

Lindeman, 2005; Newby & Davis, 2004). From a conceptual standpoint, the findings also support previous concerns about the items contained within the *Extraordinary Lifeforms* (ELF) subscale of the RPBS. The two items that were subsequently removed from the scale during analyses ("*the abominable snowman of Tibet exists*" and "*There is life on other planets*") were contained within the original three-item ELF subscale of the RPBS, supporting previous claims that these cryptozoological items do not reflect classically "paranormal" phenomena. While one item from the ELF subscale ("*The Loch Ness monster of Scotland exists*") was retained with a reasonable factor loading of .59, it is important to note that the current sample was biased towards white British participants (68.90%) who may be more familiar with this creature than samples from other cultures (as previously highlighted). Therefore, exploratory factor analysis with more varied samples is likely to produce different factor loadings for this item. Despite some similarities to previous literature, the findings from the exploratory factor analysis are in contrast with Tobacyk's (2004) original categorisation of the RPBS items and add to the disparate factor structures reported for the scale.

In recent years, psychology has been highlighted as facing a so-called "replication crisis" (see Wiggins & Christopherson, 2019; Maxwell et al., 2015, for discussions on this topic). While a review of error rates and potential bias within psychological research is beyond the scope of this work, issues surrounding successful replications are pertinent to exploratory factor analysis. As Costello and Osborne (2005) note, EFA is an error-prone statistical procedure requiring not only very large samples for optimal data, but supplementary confirmatory and reliability analyses to suitably test hypotheses or theories. Several authors have proposed varying sample-size guidance for EFA, such as Comrey and Lee's (2013) suggestion of 300 participants as being "good" for factor analyses, 500 as "very good", and 1000 as "excellent". However, such suggestions fail to consider the number of scale items or parameters estimated in the analysis (Osborne & Fitzpatrick, 2012).

Recommendations accounting for these factors suggest between five and twenty participants per scale item (Stevens, 2012), or a minimum of 10 participants per parameter estimated (Jöreskog & Sörbom, 1996). It is, therefore, important to consider the sample sizes of previous studies aiming to replicate and explore the factor structure of the RPBS. While studies (e.g., Thalbourne, 1995; Thalbourne et al., 1995) generally meet the suggested minimum sample size of five participants per scale item (5 [participants] x 26 [RPBS items] = 130 [total sample]), these fall considerably short of the 20 participant per item suggestion (520 total sample) and the 10 participants per parameter estimated (26 [RPBS items] x 7 [RPBS suggested subscales] = 182 [parameters estimated]; 182 [parameters estimated] x 10 [participants] = 1,820 [total sample]). It should be noted, however, that the seven-factor solution for the RPBS has been replicated in one large sample which exceeded the 1,820 suggested minimum sample size based on the number of parameters estimated in the analysis (N = 3,764; see Drinkwater et al., 2017). The issues raised here could be considered limitations of the present work, which had a total sample of 167 (approximately 6 participants per scale item). In addition to a smaller sample size, the current study did not perform further tests of reliability and validity of the EFA derived 24-item scale (e.g., confirmatory factor analysis, test-retest reliability analysis, internal replicability analysis).

The reliability of, and confidence in, factor analytic methods are not only limited by their sample size requirements, but by the sample-dependent person-item parameters and the lack of item characteristic assessment associated with classical test theory (Magno, 2009; DeVellis, 2006; Hambleton & Jones, 1993). As previously discussed, the RPBS (developed through the classical test theory method of factor analysis) has been identified as suffering from differential item functioning, limiting the conclusions that can be drawn from data collected with this scale. While modern test theory methods, such as Rasch analysis, still require substantial sample sizes to produce robust data, they allow for more sensitive and precise assessment of psychometric tools. Owing to the issues surrounding both the RPBS and the statistical techniques used to develop it, future work would benefit from the use of larger, more heterogeneous, samples and modern test theory methods to develop a more refined and accurate measure of paranormal beliefs within the general population.

STUDY 3.2. DEVELOPMENT OF THE PARANORMAL AND SUPERNATURAL BELIEFS SCALE USING CLASSICAL AND MODERN TEST THEORY

The aim of Study 3.2. was to address the previously highlighted issues with existing paranormal belief measures (including subscales that are often heavily culture specific or do not reflect mainstream beliefs commonly associated with the paranormal, a lack of negatively phrased items, and the potential for DIF). The study sought to address these issues by developing a scale that included phenomena widely considered to be associated with the paranormal, had less culture-bound items, combined both positively and negatively phrased items, and did not contain evidence of DIF. Therefore, the study aimed to construct a scale for measuring paranormal beliefs, examine the latent structure, and refine the scale using both CTT and MTT models. Finally, the study sought to determine the usefulness of each psychometric method by comparing the CTT and MTT analyses, and to determine which method allows for the most precise measurement of belief in the paranormal.

METHOD

Participants

An opportunistic sample of the general public (N=343) was recruited through advertisements placed on social media. These advertisements asked for participants over the age of 18 and fluent in English to complete several short questions about their beliefs in paranormal and superstitious phenomena, as well as a few short questions about themselves. Removal of incomplete responses resulted in the final sample (N=231: 83 males and 144 females, 4 unreported: Age 18–80, M=36.94, SD=14.60). Most participants were white (51.10% white British, 21.20% other white background, 06.90% White Irish) and held an undergraduate degree or higher (71.00%). Of the participants with a university education, most had a background in psychology (21.60%).

Materials

An initial collection of 29 statements was generated using adapted items from: the RPBS, the ASGS, and the SSUB, as well as four novel items. These novel items arose from discussion and examination of the RPBS, ASGS and SSUB to identify any phenomena absent from these measures, such as possessions and protection objects. Examples of the phenomena used include luck (lucky charms and bad luck), psi (sixth sense and psychics) and hauntings (Ouija boards and possession). The scale contained both positively (n=23) and negatively phrased items (n=6).

Procedure

The scale was administered as an online survey using Qualtrics Survey Software (Qualtrics, Provo, UT; see https://www.qualtrics.com). Participants were informed that the study was concerned with paranormal and superstitious belief within the general population. Respondents who agreed to take part were asked to provide their age, gender (male, female, other), ethnicity (Arabic, Asian/ Asian British, Bangladeshi, Black/Black British, Chinese, Indian, Pakistani, White British, White Irish, other Asian background, other White background, mixed background) level of education (doctoral degree, postgraduate degree, undergraduate degree, post-secondary education, secondary education, vocational) and academic discipline if they had indicated a university education (architecture, arts and humanities, business, education, law, medicine, natural sciences, philosophy, psychology, social sciences, theology, technology, other medical, other). Respondents had the option not to provide the above demographic details. Participants then completed the paranormal scale. Responses were recorded using a 7-point Likert scale (Strongly Disagree, Moderately Disagree, Slightly Disagree, Uncertain, Slightly Agree, Moderately Agree, Strongly Agree). The seven response options were numerically coded from 1 to 7 for positively worded items, and reverse coded for the negatively worded items. Following completion of the scale, participants were asked if they would be willing to complete the scale again oneweek from the date of initial completion.

Informed consent was obtained from all participants and all methods were performed in accordance with relevant guidelines and regulations. Ethical approval for the study was granted by the University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (HSET ECDA; protocol number LMS/PGR/UH04161).

Data analysis

Analyses were conducted using two models: a classical test theory (CTT) model and a modern test theory (MTT) model. Therefore, the analysis used both an exploratory factor analysis (EFA) and a rating scale model (Rasch model). The EFA allows for the identification of underlying latent constructs underpinning the scale. In other words, the EFA was used to identify emerging subcategories (or factors) across the initial collection of 29 items. Factors emerging through EFA were interpreted as distinct categories of paranormal belief. EFA was conducted using a principal components extraction method, selecting only eigenvalues greater than 1, and a direct oblimin rotation. Items with factor loadings <.50 were removed from the scale and the EFA run again until all items demonstrated acceptable factor loadings. EFA also explored group differences and answering patterns to the scale items and factors to further assess the effectiveness of the remaining scale items.

Rasch analysis was conducted to allow for a comparison between CTT and MTT methods of scale development. Owing to the polytomous nature of the data, a rating scale model (RSM; Andrich, 1978) was adopted for the Rasch analysis. Analyses first evaluated item thresholds and item characteristic curves (ICCs) for the initial collection of 29 items to assess the suitability of the 7-point Likert response format. Item fit to the model was then assessed by examining both infit (weighted) and outfit (unweighted) mean square statistics (MNSQ). Items identified for overfitting MNSQ $<.07/t<-2$) or underfitting/misfitting (MNSQ $>1.2/t>2$) were removed from the scale (Smith et al., 2008). The person-item map was then consulted to assess item difficulty, and to determine whether the remaining items meaningfully measure the ability (level of belief) of all persons. Therefore, the person-item map was used to determine whether the final scale is suitable for measuring the range of paranormal belief (from low belief/scepticism to high belief). A CTT method of confirmatory factor analysis (CFA) was used alongside the Rasch analysis to confirm the unidimensional model fit of the RSM. Finally, remaining items were tested for DIF in relation to: age, gender, ethnicity, education, or discipline.

A test–retest reliability analysis was also conducted for both the CTT and MTT scales.

RESULTS: CLASSICAL TEST THEORY

Factor analysis of the scale

An exploratory factor analysis (EFA) was conducted to investigate the latent constructs underpinning the scale. A principal components extraction method was employed and only eigenvalues greater than one were extracted. A direct oblimin rotation was used to account for the non-orthogonality of the items. Bartlett's Test of Sphericity was significant (χ^2=4975.77, p<.001) and the Kaiser–Mayer–Olkin value equalled 0.95 indicating that the data were suitable for further analysis. A four-factor solution was extracted, accounting for 64.32% of the total variance. Cronbach's Alpha was computed for each factor, with all four showing good internal consistency (α>.70). Examination of the pattern matrix revealed seven items with low item loadings (<.50), and so a second analysis was undertaken after excluding these items. The second analysis conducted on 22 items indicated a three-factor solution, accounting for 63.94% of the total variance. Inspection of the pattern matrix revealed a further two items with loadings <.50, leading to an analysis restricted to 20 of the scale items. The final analysis accounted for 65.67% of the total variance. All emergent factors demonstrated good levels of internal consistency and were conceptually distinct. Of the nine items that were removed during EFA, most were concerned with belief in psychics and those with supernatural abilities (e.g., "psychokinesis, the movement of objects through psychic powers, does exist", "tarot cards are an accurate way to see a person's past, present, and future", "astrology is a way to accurately predict the future", "mind reading is possible"). The first factor, eigenvalue 10.07, accounted for 50.34% of the variance and demonstrated excellent internal reliability (α=.95). The 14 items contained within Factor 1 concerned phenomena such as spell casting, communicating with the dead, hauntings, possession, the soul, and premonitions. As this factor contained 70% of the total scale items and covered a variety of paranormal phenomena that could be considered supernatural, Factor 1 was subsequently labelled "*Supernatural Beliefs*". The second factor had an eigenvalue of 1.87 and accounted for 9.34% of the variance. Factor 2 showed excellent internal reliability (α=.88). The factor comprised three items concerned with common superstitions centred around bad luck. Factor 2 was subsequently labelled "*Bad Luck*". The final factor, eigenvalue 1.20, accounted for 5.99% of the variance, with low to moderate internal reliability (α=.53). Factor 3 comprised three items regarding telepathy, charms, and predicting the future, and was labelled "*Psi*".

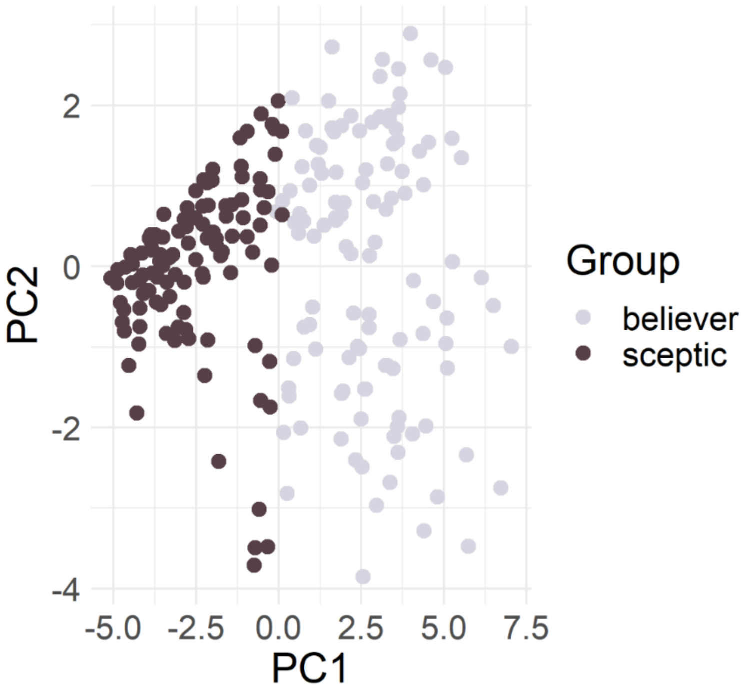
Response differences between believers and sceptics

Participants were divided into groups of 'believers' and 'sceptics' according to their mean scores (with those scoring below the overall mean of 67.30 identified as 'sceptics' and those above as 'believers'). The total sample comprised 117 (50.60%) sceptics and 114 (49.40%) believers.

Principal component analysis

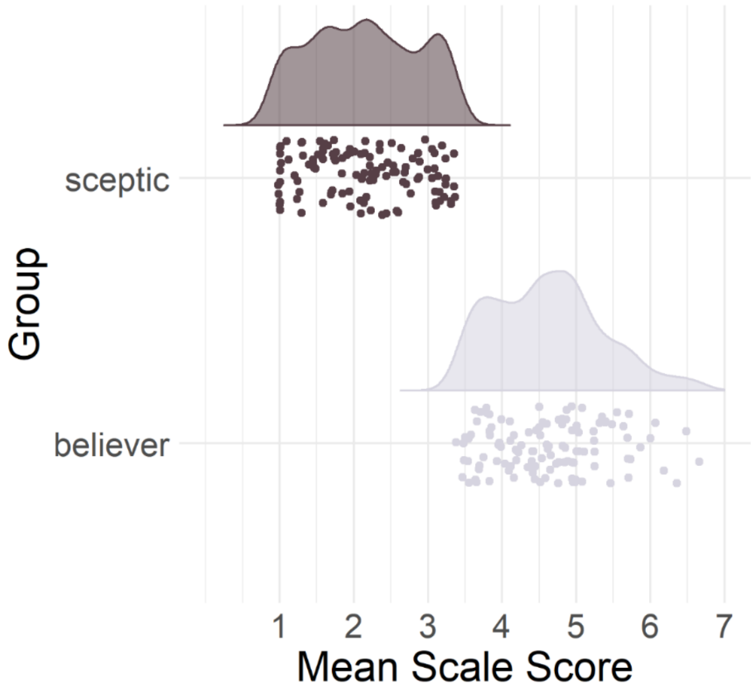
To provide a visual overview of answering patterns for the two groups, a principal component analysis (PCA) was conducted using the ggfortify (Tang et al., 2016) package in R version 4.0.2 (R Core Team, 2020). The PCA score plot (see Figure 3.3.) shows responses to all 20 items as a function of respondent group and highlights the distinct clustering of believers and sceptics, with very little overlap between the two groups. To visually represent the responses to each item on the scale for believers and sceptics, a raincloud plot was created, and the results can be seen in Figure 3.4.

Figure 3.3. PCA score plot of all responses to the paranormal scale as a function of respondent group.



Note: Figure plots participants' responses to the scale items against the two principal components that represent the largest variability among the two groups, to provide a visual indication of the separation (or lack thereof) between the groups.

Figure 3.4. Raincloud plot of mean scale scores given as a function of respondent group.



Note: Figure presents mean Likert scores (from 1-7) for all items on the scale, with individual mean scores per participant shown for each group, and a histogram showing the distribution of mean scale scores for each group.

Group answering patterns

Responses for believers and sceptics were tested for each item and factor. Table 3.6. displays the percentage agreement for each item and subsequent factor across both groups. Responses labelled "strongly disagree", "moderately disagree" and "slightly disagree" were collapsed to give an overall "disagree" score for a given item or factor. The same was done for responses labelled "strongly agree", "moderately agree" and "slightly agree" to provide an overall "agree" score. Participants' percentage of "uncertain" responses are also shown here as a function of respondent group. Percentage agreement was also calculated for participants in the upper and lower quartiles to provide a more accurate reflection of item-based differences for the most sceptical participants and those with the strongest paranormal beliefs (see Table 3.7.).

Table 3.6. Percentage agreement with factors and items as a function of respondent group.

	Disagree %		Uncertain %		Agree %	
	Believers	Sceptics	Believers	Sceptics	Believers	Sceptics
Factor 1 Total	*13*	*75*	*19*	*12*	*67*	*13*
Item 1	4	60	16	13	80	27
Item 3	8	55	18	22	74	23
Item 5	31	90	25	5	45	5
Item 7	14	71	22	20	64	9
Item 8	11	87	16	5	73	8
Item 9	13	82	9	13	78	5
Item 10	3	64	11	15	87	21
Item 12	8	72	28	17	64	11
Item 13	22	91	11	5	67	4
Item 15*	9	61	24	18	68	21
Item 16	13	68	10	7	77	26
Item 17*	13	80	31	11	56	9
Item 18	20	86	25	6	54	8
Item 20	18	91	26	7	56	3
Factor 2 Total	*64*	*94*	*12*	*2*	*23*	*4*
Item 2	61	92	15	3	25	5
Item 4	60	95	11	3	30	3
Item 6	73	95	11	1	16	4
Factor 3 Total	*30*	*69*	*20*	*6*	*49*	*25*
Item 11*	46	85	14	1	39	15
Item 14*	23	57	24	10	54	32
Item 19*	21	64	24	8	55	28

*Note: * = reverse scored items, table presents the percentage of believers and sceptics who indicated agreement, disagreement, or uncertainty for each item and each factor.*

Table 3.7. Percentage agreement with factors and items for upper and lower quartiles.

	Disagree %		Uncertain %		Agree %	
	Upper Quartile	Lower Quartile	Upper Quartile	Lower Quartile	Upper Quartile	Lower Quartile
Factor 1 Total	*6*	*91*	*11*	*4*	*83*	*4*
Item 1	3	82	3	5	93	13
Item 3	2	84	7	8	91	8
Item 5	12	98	19	0	69	2
Item 7	7	85	16	11	78	3
Item 8	2	98	3	2	95	0
Item 9	3	98	0	2	97	0
Item 10	0	92	3	3	97	5
Item 12	2	94	21	6	78	0
Item 13	12	98	12	0	76	2
Item 15*	3	76	19	13	78	11
Item 16	7	82	3	5	90	13
Item 17*	9	94	16	2	76	5
Item 18	10	97	16	2	74	2
Item 20	5	100	17	0	78	0
Factor 2 Total	*55*	*99*	*15*	*0*	*30*	*1*
Item 2	55	98	17	0	28	2
Item 4	45	100	10	0	45	0
Item 6	64	100	17	0	19	0
Factor 3 Total	*19*	*74*	*23*	*5*	*58*	*22*
Item 11*	28	82	17	2	55	16
Item 14*	12	69	24	8	64	23
Item 19*	17	69	28	5	55	26

*Note: * = reverse scored items, table presents the percentage of participants in the upper and lower quartiles who indicated agreement, disagreement, or uncertainty for each item and each factor.*

To test for differences in the two groups, items were then stacked by factor and Chi-Square analysis was conducted. Believers and sceptics differed reliably on all factors, with believers scoring significantly higher than sceptics (i.e., agreeing with more of the statements) for each of the three factors (see Table 3.8.). Examination of the group answering patterns revealed that, while most believers agreed overall with Factors 1 and 3, a higher proportion disagreed with Factor 2. Therefore, it can be said that the items in Factor 2 are less effective in separating believers and sceptics, particularly when compared to the percentage scores for Factor 1. Inspection of Table 3.8. revealed that the scores for believers and sceptics were most similar for Factor 2, with Factors 2 and 3 both displaying small effect sizes. As Factors 2 and 3 both presented limitations (both had small effect sizes, Factor 2 was less effective in separating the two groups, and Factor 3's internal reliability was below satisfactory thresholds), a final exploratory factor analysis was conducted removing the six items contained within Factors 2 and 3. The analysis used the same extraction and rotation methods as before. Bartlett's Test of Sphericity was significant (χ^2=2565.14, $p < .001$) and the Kaiser–Mayer–Olkin value equalled 0.95 indicating that the data were suitable for further analysis. A one-factor solution was extracted, accounting for 62.93% of the total variance. Cronbach's Alpha was computed for this factor, which retained the excellent internal consistency found in the earlier analysis (α=.95). Table 3.9. presents the final 14 items contained within the single factor alongside the component loadings seen in the (non-rotated) component matrix.

Table 3.8. Mean score (standard errors), χ^2, *p* values, and Cramer's V for likelihood ratio tests for groups within each factor.

	Mean (SE)				
Factor	Believers	Sceptics	χ^2	*p*	Cramer's V
1	5.07 (.10)	2.19 (.11)	1330.63	<.001	.45
2	2.82 (.12)	1.41 (.07)	93.24	<.001	.26
3	4.32 (.11)	2.64 (.14)	105.83	<.001	.28

Demographic differences

Owing to the somewhat mixed research suggesting a correlation between paranormal beliefs, academic discipline and aspects of thinking, responses to the paranormal scale were compared for those with and without higher education backgrounds; and between those from science and non-science academic disciplines. Most participants held an undergraduate degree or higher (n=164), while less than half held post-secondary qualifications or lower (n=67). Participants with university degrees had lower total paranormal scores (M=46.20, *SD*=22.89) than participants without university degrees (*M*=61.34, *SD*=22.08). The difference in scores between the two education groups was significant [$t(126.78)=-4.68$, $p < .001$]. Of the participants with degree qualifications, most were from science-based disciplines including psychology, natural sciences, technology, and other medical backgrounds (n=83), while the rest included social sciences, education, business, philosophy, theology, art and humanities, law, and architecture (n=57). As 24 participants did not disclose their discipline, the following analyses were conducted on 140 participants. Those from science-based disciplines demonstrated lower paranormal scores (*M*=40.02, *SD*=21.28) compared to those with art-based degrees (*M*=54.77, *SD*=22.24), and the difference in scores between the two discipline groups was significant [$t(116.99)=3.92$, $p<.001$].

Table 3.9. Single-factor scale with corresponding Cronbach's Alpha (α) score and component loadings.

Factor	α	Items (loading scores)
1 Supernatural Beliefs	.95	1 The soul continues to exist after a person has died (.76) 2 Your mind or soul can leave your body (.77) 3 It is possible to cast spells on persons using formulas and incantations (.80) 4 It is possible to be reincarnated (.74) 5 Some people with psychic abilities can accurately see the future (.86) 6 It is possible to communicate with the dead (.86) 7 Buildings can be haunted by spirits or other supernatural entities (.87) 8 Some psychics have helped find the bodies of murder victims through paranormal means (.85) 9 A person's star sign can have a direct influence on their personality (.76) 10* Reports of an apparent sixth sense are generally based on fantasies (.72) 11 Having a dream that comes true is not just a coincidence (.71) 12* Communicating with spirits or other supernatural entities through a Ouija board is not possible (.75) 13 It is possible to become possessed by an evil supernatural entity (.81) 14 It is possible to protect one's home from spirits using protection objects and herbs (.83)

*Note: * = reverse scored items*

Test-retest reliability

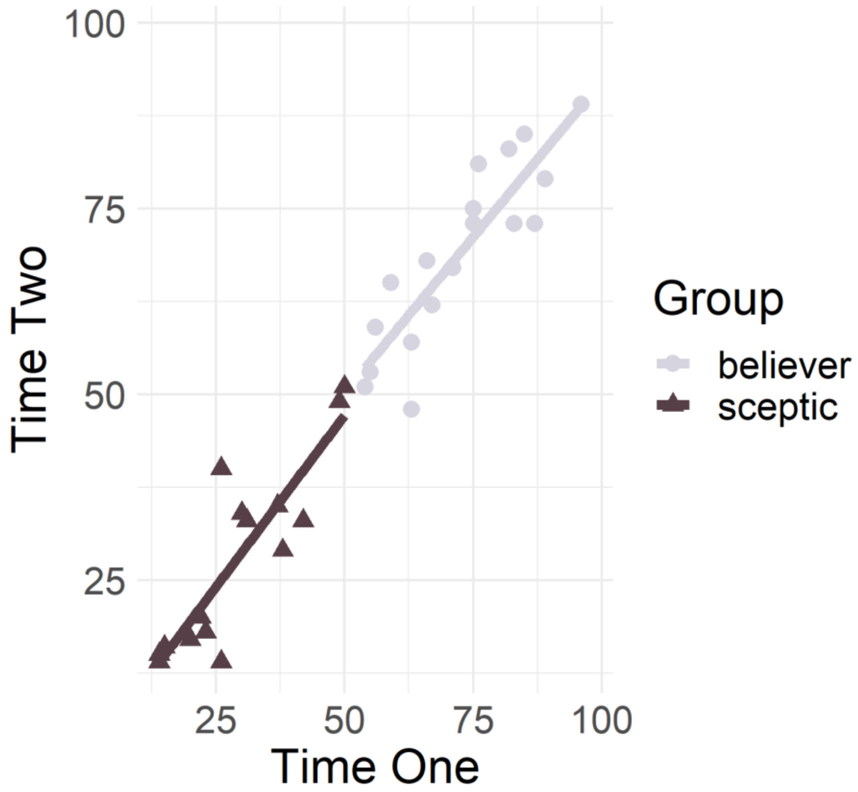
Sample and procedure

A follow-up study was conducted to assess the test–retest reliability of the newly developed scale. Of the original sample of 231 participants, 37 (16% of the original sample) agreed to complete the scale a second time, oneweek after their initial participation. The retest sample consisted of 21 males (56.80%) and 16 females (43.20%), aged between 18 and 73 (M=41.51, SD=16.61). In contrast to the original sample, this sample had a higher percentage of male participants and a higher mean age. The difference in gender between the original participant group and the retest group was significant (χ^2=5.433, *p*=.02). However, the difference in age between the two groups was not significant [*t*(262)=-1.77, *p*=.078]. Nineteen respondents were identified as 'sceptics' (51.35%) and 18 as 'believers' (48.65%), according to their mean scores on the 14-item scale at time one (with those scoring below the overall mean of 50.59 identified as 'sceptics' and those above as 'believers'). The questionnaire completed by participants comprised the original collection of 29 statements and used the same 7-point Likert response format (Strongly Disagree, Moderately Disagree, Slightly Disagree, Uncertain, Slightly Agree, Moderately Agree, Strongly Agree). Responses were numerically coded as before. The scale was administered again as an online survey using Qualtrics Survey Software.

Retest analysis

Retest analyses were conducted on the final 14-item scale. Pearson's correlations revealed a strong test–retest reliability for the scale [*r*(35)=0.98, $p < .001$], as well as for both believers [*r*(15)=0.88, $p < .001$] and sceptics [*r*(18)=0.90, $p < .001$]. A scatterplot of the scores for believers and sceptics at time one and time two can be found in Figure 3.5.

Figure 3.5. Test-retest reliability analysis as a function of respondent group.



Note: Pearson's correlations between participants' individual total scores at time one and time two shown for each group.

RESULTS: MODERN TEST THEORY

The MTT analyses presented in the following sections were conducted using a Rasch rating scale model (RSM) using the eRm (Mair et al., 2021; Mair & Hatzinger, 2007) package in R version 4.0.2 (R Core Team, 2020).

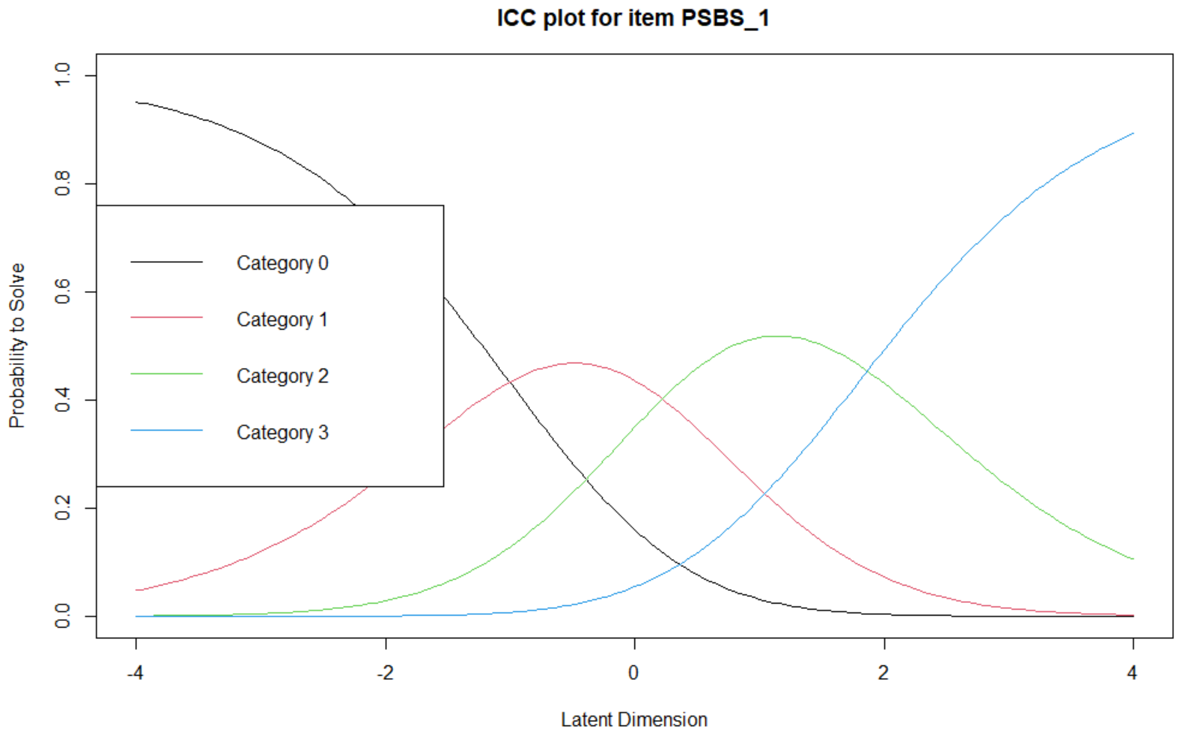
Response categories

MTT analyses first focused on evaluating the effectiveness of the 7-point Likert rating scale. As it is difficult to be certain of the exact way the sample will use the rating scale, investigation is necessary to verify or improve the functioning of the rating scale categories (Linacre, 2002A). To evaluate the response category use of the sample, threshold parameters of each category were examined for each of the original 29 items. These thresholds identify and define the boundaries between each response category and should, therefore, increase monotonically, such that the probability to solve (the probability that a response category is selected) reflects the respondent's position along the latent trait of interest. Consequently, participants with higher levels of paranormal beliefs should be more likely to endorse higher response categories. For the Rasch analyses, responses are shifted such that the lowest category (strongly disagree) is 0.

Analysis of the 7-point rating scale revealed that threshold parameters failed to increase monotonically, therefore indicating evidence of step disordering. Step disordering, occurring when threshold parameters fail to increase monotonically, indicates that certain response categories have a low probability of being observed (Linacre, 2002A), meaning that the sample are less likely to use these response categories. The lack of ordered increase occurred at Category 2 (somewhat disagree). Examination of the item category curves (ICCs) indicated that Category 2 had the lowest probability of observance and was therefore never more likely to be observed than any other category. Put more simply, regardless of an individual's level of belief in paranormal phenomena, the probability of choosing "somewhat disagree" is never the most likely. Similarly, Category 1 (moderately disagree) also had a low probability of observance and at no point was this category most likely to be observed.

To begin to improve the functioning of response categories, responses were recoded such that the "moderately disagree" and "somewhat disagree" categories were collapsed, as were the "moderately agree" and "somewhat agree" categories. Tis gave a revised 5-point scoring method (0=strongly disagree, 1=disagree, 2=uncertain, 3=agree, 4=strongly agree). However, this revised scoring method failed to rectify step disordering. Examination of the ICCs revealed that the boundaries between Categories 1 and 2 (disagree and uncertain) were very narrow and suggested that the sample did not clearly differentiate between these two categories. Therefore, a final recoding took place such that the "disagree" and "uncertain" categories were collapsed, giving a final revised4-point scoring method (0 = strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree). When this final scoring method was used, the four categories increased monotonically, with the desired appearance of the range of peaks for each category appearing in the ICCs for each item. An example of the ICC for item 1 is shown in Figure 3.6.

Figure 3.6. Item characteristic curve for item 1 using the 4-point scoring method.



Note: Curves represent the probability of selecting a category along the latent trait. Category 0 = "strongly disagree", Category 1 = "disagree", Category 2 = "agree", Category 3 = "strongly agree".

Item fit

Mean square statistics (MNSQ) were computed to determine item fit to the model (i.e., how well each item contributes to defining a single unidimensional construct). The MNSQ statistics indicate the amount of distortion of the scale, where high MNSQ values indicate unpredictability and a lack of construct similarity with other scale items (underfitting), and low values indicate item redundancy and less variation in the observed data compared to the variation that was modelled (overfitting; Linacre, 2002A). Two MNSQ statistics were used to assess item fit: infit (weighted) and outfit (unweighted) statistics. Subsequent analyses used an accepted range of fit of 0.7 to 1.2 (Smith et al., 2008) to identify items with poor model fit. Therefore, items with MNSQ values < 0.7 were identified as overfitting the model, and MNSQ values > 1.2 were identified as underfitting the model. When assessing item fit to the model, infit and outfit t-statistics were also examined where t-values < -2 were identified as overfitting and t-values > 2 were identified as underfitting. However, it has been suggested that infit and outfit MNSQ values are relatively insensitive to sample size variation in polytomous data, while the t-statistics vary considerably with sample size. Therefore, it has been recommended that infit and outfit t-statistics are interpreted with caution when determining item fit to the model for large samples and polytomous data (Smith et al., 2008). As such, items would be removed from the scale if they demonstrated both infit and outfit MNSQ values that were overfitting or underfitting the model. In cases where items were only identified on one of the MNSQ values (infit *or* outfit), t-statistics were consulted to verify item misfit. Based on the MNSQ values of the 29 items, a total of 7 items (4, 10, 12, 13, 15, 28 and 29) were identified for overfitting and a further 8 items (1, 2, 5, 8, 14, 17, 23 and 27) were identified for underfitting. Subsequently, these 15 items were removed from the scale and the analysis was conducted again on the remaining 14 items. A final item (7) was identified for overfitting the model and was removed from the scale. Analysis of the final 13 items revealed infit and outfit statistics within the specified ranges. While item 11 produced an infit t-statistic of -2.2, the infit and outfit MNSQ values were within the specified range (0.81 and 0.83, respectively) as was the outfit t-statistic (-1.76). Considering these other statistics and given that the infit t-statistic of item 11 was very close to -2, it was determined that the item demonstrated reasonable fit to the model and that there was not sufficient evidence to remove the item from the final scale. Table 3.10.

shows the final MNSQ statistics for the remaining items, along with the corresponding item difficulty statistics.

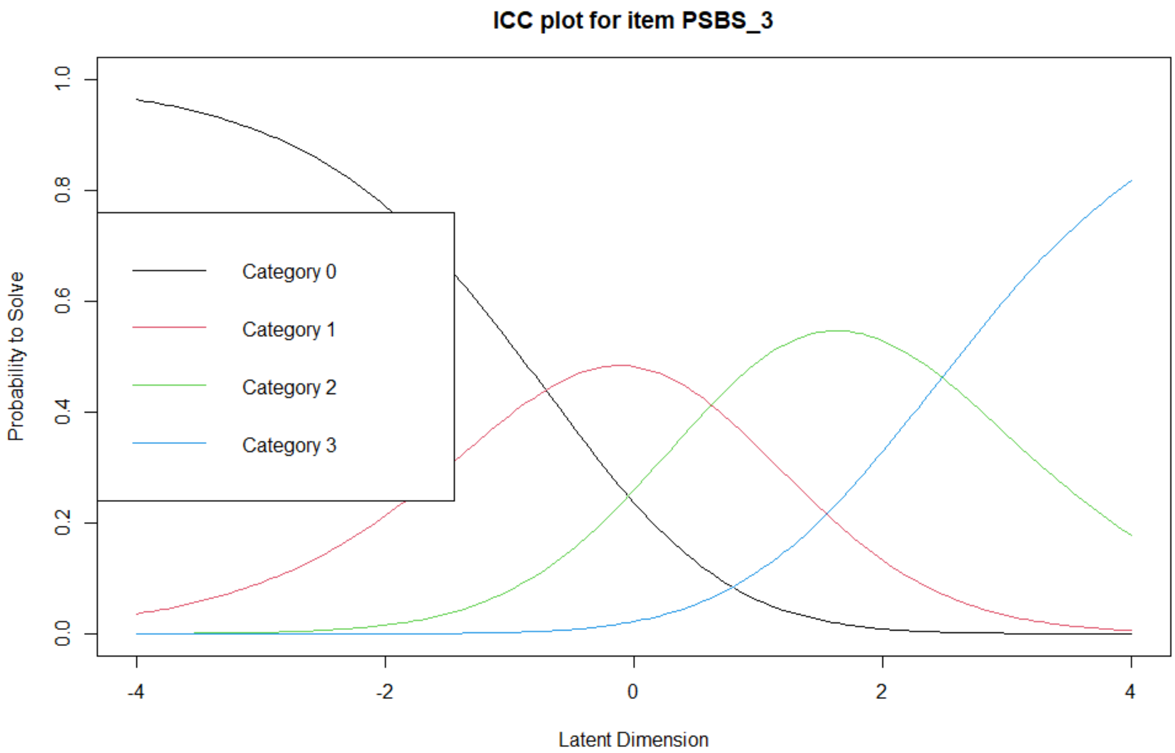
Table 3.10. Parameter values for the remaining 13 items (in order of item difficulty).

Item	Outfit MSQ	Infit MSQ	Difficulty
6 If you break a mirror, you will have bad luck.	1.002	1.112	2.547
18 Fairies and similar beings are real.	1.001	0.866	2.432
19* Fortune tellers' predictions are typically based on guesswork.	1.046	0.823	2.026
16 A person's star sign can have a direct influence on their personality.	0.971	0.993	1.663
21 Some health conditions can be treated with psychic healing.	0.986	0.937	1.587
26 It is possible to become possessed by an evil supernatural entity.	1.018	0.991	1.540
25* Communicating with spirits or other supernatural entities through a Ouija board is not possible.	1.199	1.170	1.475
11 Mind reading is possible.	0.832	0.809	1.401
9 It is possible to be reincarnated.	1.049	0.974	1.318
22 In some cultures, shamans or "witch doctors" exercise powers we cannot explain.	0.892	0.862	1.199
20* Reports of an apparent sixth sense are generally based on fantasies.	0.829	0.833	0.829
3 Your mind or soul can leave your body.	1.074	1.058	0.793
24 Having a dream that comes true is not just a coincidence.	0.832	0.837	0.766

*Note: * = reverse scored items*

Owing to the substantial change in the number of scale items, thresholds for the 4-point response scale were consulted to verify the functioning of the new rating scale for the remaining 13 items. The analysis demonstrated that the thresholds of the four categories increased monotonically for all remaining items. An example of the ICC for item 3 is shown in Figure 3.7., which again shows the desired range of peaks. ICC curves for the remaining items can be found in Appendix K.

Figure 3.7. Item characteristic curve for item 3 in the reduced scale using the 4-point scoring method.

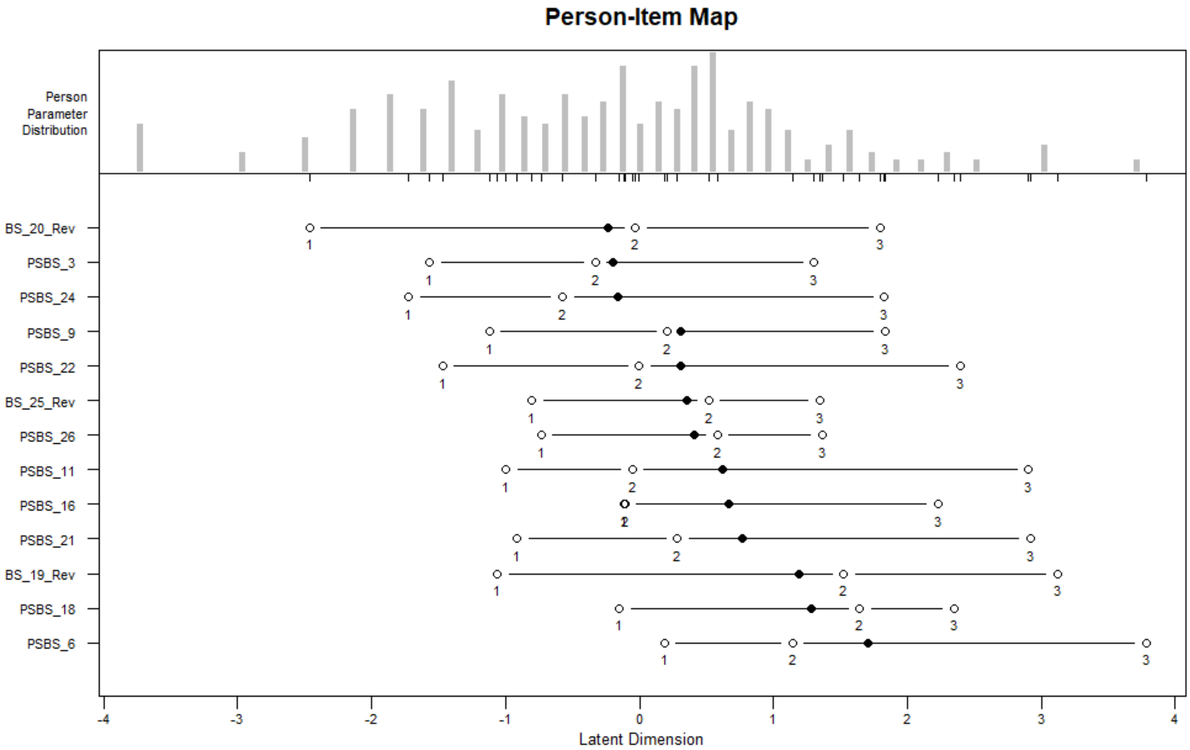


Note: Curves represent the probability of selecting a category along the latent trait. Category 0 = "strongly disagree", Category 1 = "disagree", Category 2 = "agree", Category 3 = "strongly agree".

Item difficulty

The final RSM analysis conducted using the eRm package (Mair et al., 2021; Mair & Hatzinger, 2007) sought to estimate the person trait and item difficulty parameters. In other words, the following analysis aimed to determine whether the difficulty of the remaining items was appropriate for the sample. To meaningfully measure the ability (level of paranormal belief) of all persons, items should be located along the length of the latent dimension. The person-item map shown in Figure 3.8. displays both the person traits (in the upper panel) and item difficulties (lower panel) along the same latent dimension. As shown, the category thresholds of most of the 13 items cover a low-to-high range of paranormal belief well. However, item difficulty locations (identified in Figure 1.8. as solid circles) cluster towards the right side of the latent dimension. Therefore, the items have a higher probability of differentiating between individuals with higher levels of paranormal beliefs. For example, item 6 ("if you break a mirror, you will have bad luck") shows the highest item difficulty meaning that participants with higher levels of paranormal beliefs are more likely to agree with this item.

Figure 3.8. Person-item map for the 13-item scale.



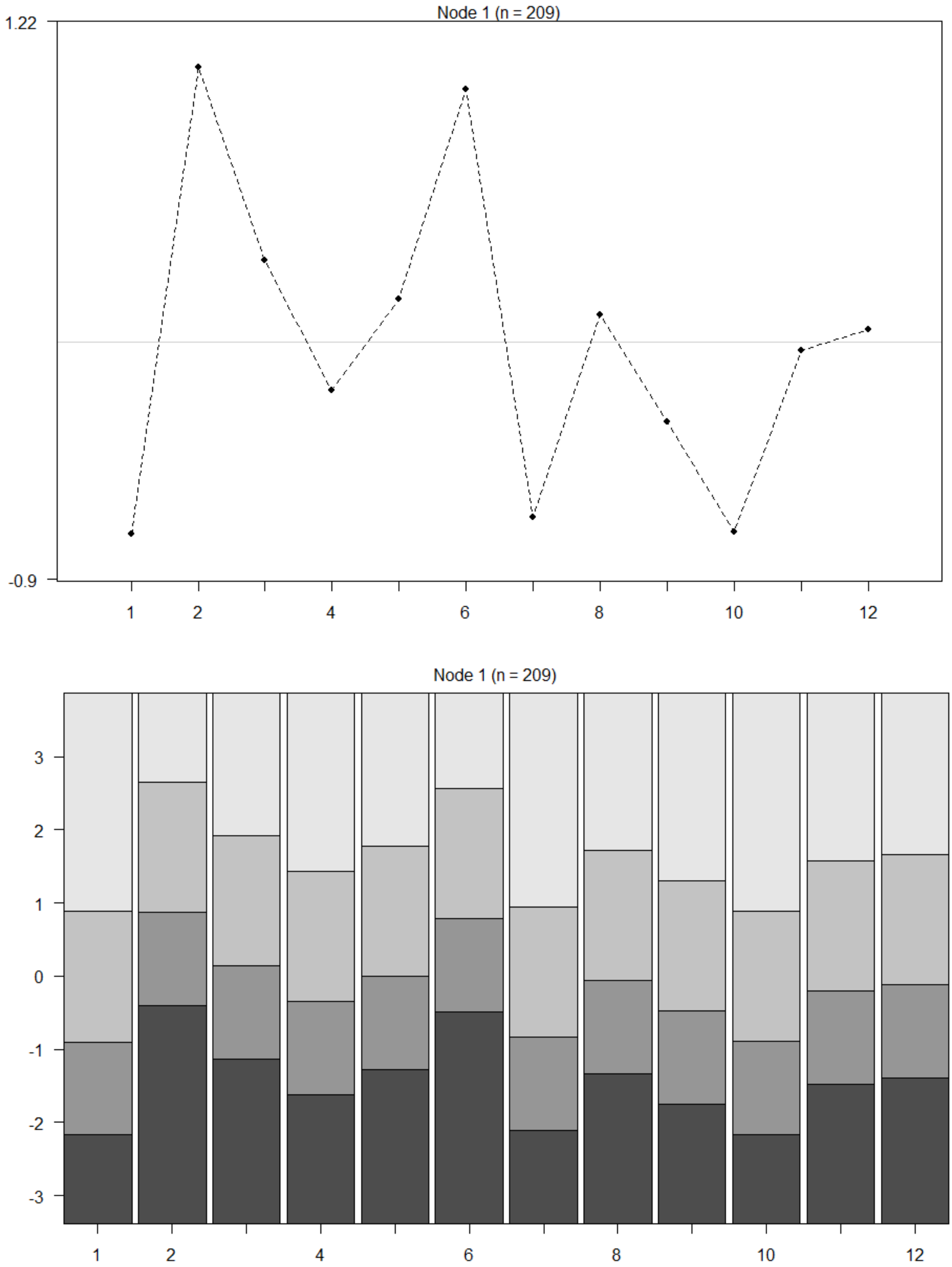
Note: Figure displays the location of person traits and item difficulties along the same latent dimension (paranormal belief). The person traits are located on the scale from left (low belief) to right (high belief). Locations of item difficulties are presented as solid circles, and thresholds of adjacent category locations are presented as open circles. The item parameters are located on the scale from least difficult (left) to most difficult (right).

Differential item functioning

Differential item functioning (DIF) analysis was conducted using rating scale trees within the psychotree (Komboz et al., 2018; Strobl et al., 2015) package in R version 4.0.2 (R

Core Team, 2020). Before this analysis was conducted, data for 8 participants who chose not to disclose demographic information were removed. Data was also removed for participants scoring only in either the highest or lowest categories (i.e., participants responding "strongly disagree" to all 13 items, or "strongly agree" to all items") as these responses do not provide information relating to item difficulty and therefore do not contribute to the Rasch model. Consequently, data for 14 participants (all of whom scored in the lowest categories) were removed. In total, 22 participants were removed and the DIF analysis was conducted on a reduced sample of 209 participants. If none of the scale items show evidence of DIF, then the analysis should produce a tree with only a single node, supporting a unidimensional Rasch model for the data (Strobl et al., 2021). However, if the Rasch tree shows at least one split and identifies more than a single node containing the entire sample, then DIF is present. These nodes would highlight item response differences between the identified groups. An advantage of using the Rasch tree method for identifying DIF is that DIF can be detected between groups of participants created by more than one covariate (e.g., females under 34), and these groups do not need to be pre-specified prior to analysis. As such, the Rasch tree method searches for the value corresponding to the strongest parameter change and splits the sample at the value identified (Strobl et al., 2021). The DIF analysis was conducted for five covariates: age, gender, ethnicity, education, and discipline. Analysis produced a tree with a single node, and therefore no DIF was present in the scale for any of the covariates. The single-node tree can be seen in Figure 3.9.

Figure 3.9. Single node Rasch tree.



Note: Figure shows difficulty values for the scale items across the total sample and no differential functioning for the items across any of the covariates.

Confirmatory factor analysis

As a final test of the unidimensionality of the scale, a confirmatory factor analysis (CFA) was conducted using the lavaan (Rosseel, 2012) package in R version 4.0.2 (R Core Team, 2020). To determine the strength of model fit, four main fit indices were used: comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardised root mean square residual (SRMR). For both the CFI and TLI, a value of .90 or above would indicate acceptable fit and a value of .95 or above would indicate very good model fit. For the RMSEA, a value of .05 or below would indicate close model fit, with a value of .08 indicating acceptable fit. The accompanying *p*-value for the RMSEA statistic should also be greater than the standardised value of .05 for close model fit. Finally, an SRMR value of .05 or below would indicate a well-fitting model. Overall, the model demonstrated good fit, and supported the use of a unidimensional Rasch model for the data. Complete fit statistics can be seen in Table 3.11.

Table 3.11. Fit statistics for confirmatory factor analysis of the final 13-item scale.

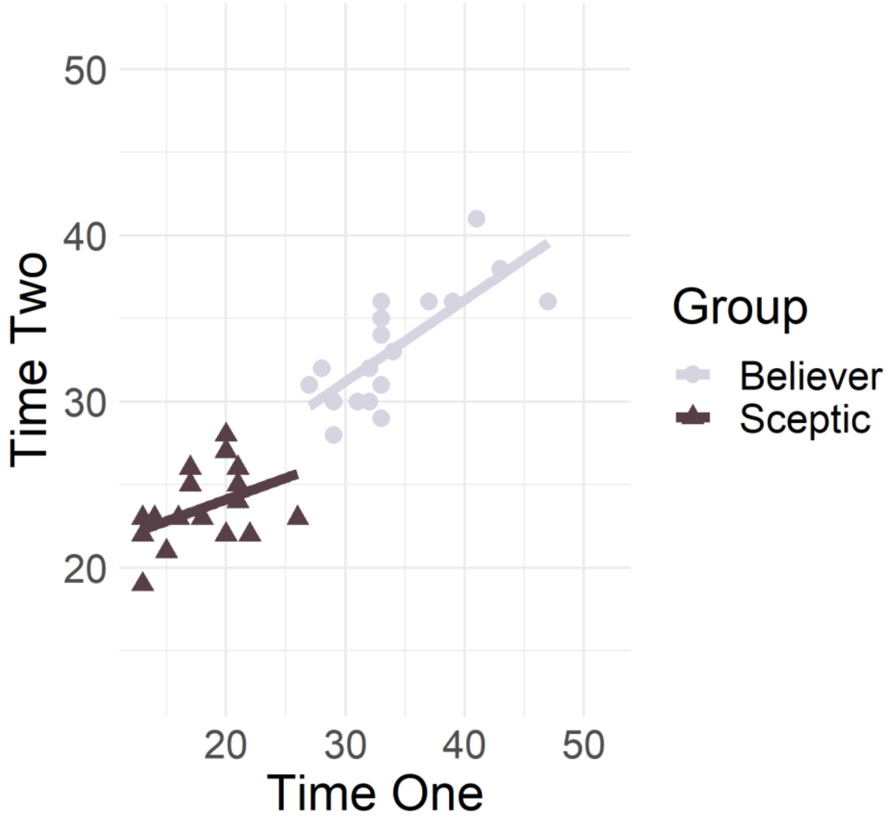
CFI	TLI	RMSEA	*p*	SRMR
.936	.923	.074	.007	.045

Note: CFI = comparative fit index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, SRMR = standardised root mean square residual.

Rasch test-retest reliability

The sample for the test-retest reliability analysis was the same as that described in the EFA analysis. While participants were divided into believers and sceptics based on their mean scores for the 13-item Rasch scale at time one (with those scoring below the overall mean of 26.94 identified as 'sceptics' and those above as 'believers'), the analysis retained the original split seen in the EFA analysis of 19 sceptics and 18 believers. Pearson's correlations revealed a strong test-retest reliability for the scale ($r(35) = .92$, $p < .001$), and for believers ($r(16) = .75$, $p < .001$). However, the retest correlation was not significant for sceptics ($r(17) = .45$, $p = .051$). A scatterplot of the scores for believers and sceptics at time one and time two can be found in Figure 3.10. Cronbach's Alpha computed for this final scale, indicated an excellent internal reliability ($\alpha = .91$).

Figure 3.10. Test-retest reliability analysis for the Rasch scale as a function of respondent group.



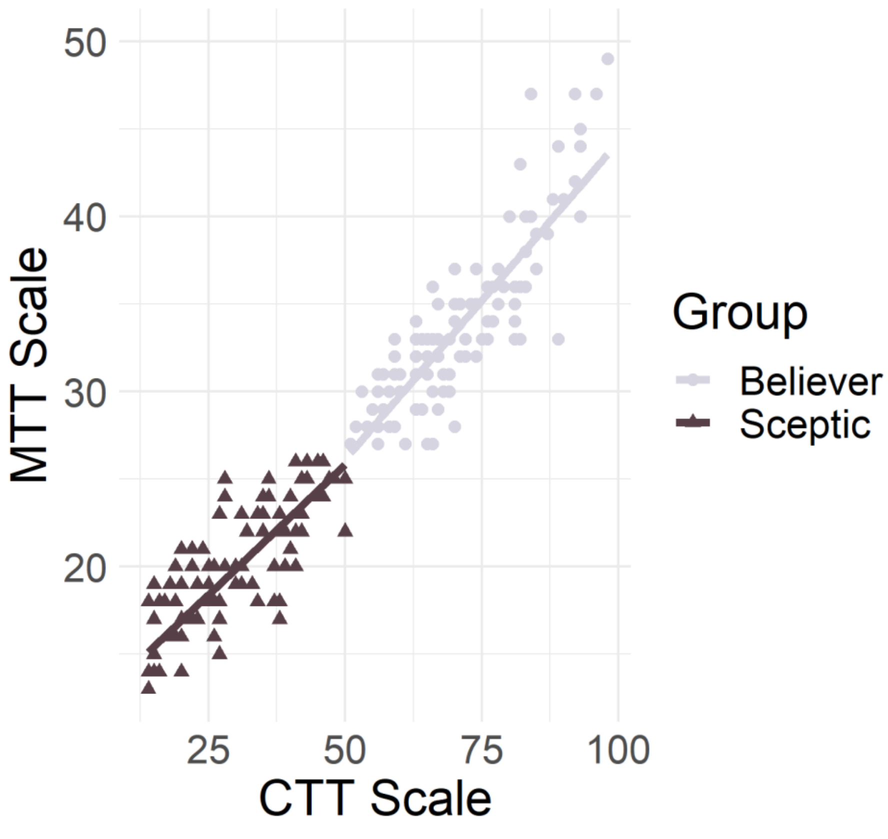
Note: Pearson's correlations between participants' individual total scores at time one and time two shown for each group.

Correlations between scales

To compare the performance of the CTT and MTT derived scales, a final correlational analysis was conducted comparing respondents' total scores on each scale. The analysis only included respondents who were identified as 'sceptics' or 'believers' by both scales.

Therefore, 17 respondents were removed from the analysis owing to the scales placing them in different groups, and the final analysis was conducted on a reduced sample of 214. Of the reduced sample, 102 respondents were identified as 'sceptics' (47.66%) and 112 as 'believers' (52.34%). Pearson's correlations revealed a strong correlation between the scales for the total sample ($r(212) = .96, p < .001$), as well as for both believers ($r(110) = .86, p < .001$) and sceptics ($r(100) = .82, p < .001$). A scatterplot of the scores for believers and sceptics at time one and time two can be found in Figure 3.11.

Figure 3.11. Correlations between respondents' individual total scores.



Note: Pearson's correlations between respondents' total scores on the classical test theory and modern test theory scales, as a function of respondent group.

DISCUSSION OF STUDY 3.2.

With a view to developing a new measure of belief in paranormal phenomena, two methods of scale development were compared. The first approach was based on the procedures of classical test theory (CTT), with a particular emphasis on exploratory factor analysis. The second approach used modern test theory (MTT) based on Rasch analysis for polytomous data. The CTT method reduced the initial collection of 29 items to a 14-item scale, describing paranormal belief on a single dimension: Supernatural Beliefs. MTT analyses produced a final collection of 13 items measured with a reduced 4-point scoring method. The final MTT derived scale is put forward as a new self-report measure of paranormal beliefs, referred to as the 'Paranormal and Supernatural Beliefs Scale' (PSBS; see Appendix L).

Several similarities can be seen between the CTT and MTT derived scales. First, both scales support a unidimensional measure of belief in paranormal phenomena. In the CTT analyses, Factor 2 (Bad Luck) initially demonstrated an excellent internal reliability. However, examination of the group answering patterns presented interesting findings, with over half of the believers' responses to these items falling under the "disagree" category. The high "disagree" scores seen for believers in Factor 2 suggest that bad luck may not be diagnostic of belief in more general paranormal phenomena, as the factor was less effective in separating believers and sceptics. For this reason, the three items contained within Factor 2 were removed from the CTT scale. The three items contained within Factor 3 (Psi) were also removed from the CTT scale as the factor did not meet satisfactory thresholds (which may be attributed to the fact that all items within this factor were negatively phrased (Taber, 2018; Roszkowski & Soven, 2010; Schriesheim et al., 1991). When initial analyses indicated three distinct categories of belief, the Supernatural Beliefs factor explained the most variance and included 70% of the total scale items. This factor was retained as the only factor for the 14-item CTT scale (α = .95), and encompassed many phenomena considered to be paranormal or supernatural (French & Stone, 2013, pp. 6-9; Irwin, 2009) suggesting that belief in the paranormal may be best characterised by a single overarching factor that is equally understood by both paranormal believers and sceptics. This provides further support for the removal of Factors 2 and 3 from the CTT scale which, while both having their own strengths and weaknesses, may represent categories of beliefs that are separable from paranormal beliefs. Item infit and outfit mean square (MNSQ) statistics (as well as differential item functioning analysis) produced through MTT analyses also indicated that the data supported a unidimensional structure, providing further support for the idea that belief in the paranormal may be best represented by a single dimension. As previous work has suggested a combination of CTT and MTT techniques for psychometric assessment (Kline, 2005), confirmatory factor analysis and reliability analysis (Cronbach's alpha) were also computed to assess the functioning of the MTT scale items as a complete unit. These findings again supported the unidimensional structure of the scale and indicated an excellent internal reliability ($\alpha = .91$). In addition to high internal reliabilities, both scales demonstrated strong test-retest reliability correlations (.98 for the CTT scale and .92 for the MTT scale). However, examination of the retest statistics for each group (believers and sceptics) revealed differences between the two scales. While the CTT and MTT scales both demonstrated good retest correlations for believers (.88 and .75 respectively, *p*s $< .001$), the retest correlation for sceptics was not significant in the MTT scale ($r(17) = .45$, $p = .051$) compared to the CTT scale ($r(17) = .90$, $p < .001$). The difference in these scores can be explained using the person-item map produced during MTT analyses, which suggested that the item within the MTT scale have a lower probability of differentiating between individuals with lower levels of paranormal beliefs. Similar differences were not able to be established through CTT analyses. To the authors' knowledge this is the first presentation of separate retest scores for believers and sceptics. Comparison of the performance of both scales revealed strong correlations between respondents' total scores on the CTT and MTT derived scales in the total sample ($r = .96$), and for believers ($r = .86$) and sceptics ($r = .82$)

separately. A final similarity between the two scales can be seen in their item content, as both scales shared 7 common items (approximately half of the total scale content).

Despite the strengths of the CTT scale, and its similarities to the MTT scale, the results of the study provide strong evidence to support preference of the MTT derived scale. First, MTT analyses allowed for investigation and refinement of the 7-point Likert scale. The results indicated that respondents did not require so many response options, and supported removal of three categories leading to a final 4-point scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). Categories 1 and 2 of the original Likert scale (moderately disagree and somewhat disagree), both had low probabilities of observance and were subsequently collapsed into a single category (as were the moderately agree and somewhat agree categories). The "uncertain" category was also found to be inadequate in representing participants' responses, with results suggesting that this category may be poorly defined with respondents not clearly differentiating between this category and the "disagree" category. A 7-point Likert scale was initially selected for the scale as it was thought that the large number of response options would produce a more precise index of respondents' level of agreement. However, these findings suggest that the response options provided in the original 7-point scale did not represent differentiable levels of belief intensity (as is indicated by a monotonic increase of category thresholds). Additionally, MTT analyses permitted an assessment of differential item functioning (DIF). Using the Rasch tree method for identifying DIF within the MTT scale, analysis focused on five covariates (age, gender, ethnicity, education, and discipline) to determine whether these, or some combination of these, influenced participants' responses to the scale. Examination of the tree revealed a single node, with no DIF identified for any of the covariates. Therefore, while the MTT scale can be described as a valid measure of belief in paranormal phenomena, it is difficult to be certain that the CTT derived scale does not suffer from DIF. As mentioned above, MTT analyses also allowed for examination of item difficulty, with results indicating that items had a higher probability of differentiating between respondents with moderate-high levels of paranormal beliefs. This information is particularly useful for future research looking to utilise the scale to examine group differences within paranormal beliefs. The following comparisons focus on the final PSBS developed through MTT analyses.

Several important differences can be noted when comparing the PSBS to the three most frequently employed measures of paranormal belief. The unidimensional structure of the PSBS is far simpler than the 7-factor RPBS, with the content of many RPBS factors (such as those within Witchcraft, Spiritualism and Precognition) appearing in the PSBS. The appropriateness of this solution accords with previous research suggesting that a larger array of factors may not provide the most prudent account of paranormal belief (French & Stone, 2013, pp. 13-14), particularly as the RPBS has an insufficient number of items to adequately sample seven distinct dimensions of paranormal belief. Such criticisms may explain why a range of studies have failed to replicate the original factor structure of the RPBS, finding smaller factor structures ranging between one and six to be more suitable (French & Stone, 2013, pp. 13-14). Despite this, most of these replication studies have suggested paranormal belief to be a multidimensional construct, which contradicts the findings from the present work. While the structure of the PSBS is more comparable to that of the ASGS (but still differs in terms of dimensionality of belief), the range of items contained within the PSBS is much broader as its focus is not confined to parapsychological phenomena such as extrasensory perception and psychokinesis, though it does include several psi-related items.

The item content of the PSBS also differs considerably from the existing scales in that the final scale presents three negatively phrased items, and contains few cryptozoological, religious, or culturally-specific items. By reducing the number of potentially problematic items and ensuring a blend of positive and negative items, the PSBS reduces the risk of biases introduced by participant response patterns and cultural differences, which have been highlighted as issues for older measures. While cultural differences are often present in paranormal beliefs (Maraldi & Farias, 2020), and consequently some PSBS items have seen cultural influence, the PSBS has a reduced number of culture-bound items compared to previous scales such as the RPBS. Therefore, the PSBS may be a stronger candidate for a universal measure of paranormal belief. A further strength of the PSBS seen particularly when compared to the RPBS, is that that the scale is not affected by certain subgroup characteristics, including respondents' age gender, ethnicity, level of education, or academic discipline. DIF analysis indicated that the PSBS is a reliable unidimensional scale that can be used to explain data from all respondents. The results seen for the DIF analysis are worth comparing to the RPBS, which contains items that are particularly sensitive to age and gender differences (Lange et al., 2000), as they suggest that the items within the PSBS have a universal application for respondents regardless of the highlighted subgroups.

Finally, there are a few limitations of the present study which should be noted. First, many of the participants involved in the study were young, well-educated, white females. While analyses confirmed that age, gender, ethnic and educational differences (including academic discipline) do not influence item functioning, further research could explore the psychometric properties of the PSBS with more varied samples and across a diverse range of cultures. As the sample size used for the test-retest reliability analysis was relatively small in the present study ($N = 37$), further research could also conduct additional analyses to confirm the stability and precision of PSBS scores across time. Furthermore, although the PSBS focuses on many phenomena that might have a universal application in practice (e.g., communication with spirits), it does present some specific examples that may be more prominent in Western cultures (e.g., Ouija boards). Finally, MTT analyses expressed that the scale is good at measuring moderate-high levels of paranormal beliefs, and so operates sufficiently for the purpose of identifying individuals with increased levels of paranormal beliefs. However, additional items that tap specifically into low levels of paranormal beliefs may be beneficial to add to the scale in future revisions to accurately capture the complete range of beliefs.

CHAPTER 3 GENERAL DISCUSSION

Measuring belief in the paranormal suffers the same limitations as the measurement of any latent construct, namely the validity and reliability of the psychometric assessment tool. As previously discussed, existing measures of paranormal beliefs, while widely used, have been criticised in relation to their validity and internal reliability. The results of Study 3.1. support these criticisms, failing to replicate the factor structure of the most widely used paranormal belief scale (RPBS) through confirmatory factor analysis. Exploratory factor analysis of the RPBS found a smaller, four-factor solution to be more appropriate than the original seven-factor solution, suggesting poor internal reliability of Tobacyk's (2004) 26-item RPBS. These findings, however, may be due in part to the small sample size in Study 3.1. which failed to account for the number of scale items or parameter estimates involved in the analyses. The existing measures of paranormal beliefs have also been criticised for the statistical methods used to develop them. In particular, the classical test theory method of exploratory factor analysis, while simple to conduct and useful for exploring trends within real test data, cannot rigorously assess item characteristics, functionality of response scales, or differential item functioning in the same way that modern test theory methods (such as Rasch analysis) can. The strengths and limitations of these scale development methods can be seen in Study 3.2. Both the exploratory factor analysis (EFA) and the Rasch analysis supported a unidimensional structure for the scale, and the EFA-derived scale produced excellent internal and external reliability statistics (Cronbach's Alpha and Pearson's test-retest correlation). While the EFA provided an indication of the latent construct(s) underpinning the scale, this method was unable to provide detailed statistics relating to item characteristics (e.g., item difficulty), differential item functioning, and the functionality of the Likert response scale. The latter is particularly important, as the Rasch analysis conducted in Study 3.2. reduced the rating scale substantially from 7-points to 4-points. Without this analysis (i.e., completing only an exploratory factor analysis to develop the scale, like in the development of the RPBS), information about participants' use of the rating scale would remain unknown, thereby increasing the risk of inaccurate construct measurement. This is a potential limitation of existing paranormal belief measures (e.g., RPBS, ASGS) which did not assess the functionality of their rating scales.

While the statistical techniques used to develop psychometric assessment tools are important, measurement of latent constructs (such as paranormal beliefs) also relies heavily on clear concept definition. A limitation of this field of research is that there is no universally agreed definition of what constitutes paranormal phenomena (although many defer to Broad's, 1949, definition that phenomena can be considered to be paranormal when they defy the basic limiting principals of current scientific understanding). Consequently, existing paranormal belief scales have been criticised for including phenomena that might constitute different belief clusters. In particular, the RPBS has been criticised for including items relating to religious beliefs, which (while sharing some overlap with paranormal beliefs) have been argued to stem from separate cultural spheres (see Baker & Draper, 2010; Stark & Bainbridge, 1980). Some support for this idea can be seen in the results of the exploratory factor analysis in Study 3.1., with items relating to religious phenomena grouped together in a single factor, suggesting that these items represent a distinct belief type separate from more general paranormal beliefs. The findings from Study 3.1. also support criticisms relating to the inclusion of cryptozoological items in paranormal belief scales, with two of the three "*Extraordinary Life Forms*" items removed during the exploratory factor analysis owing to poor factor loadings. These items have been criticised for reflecting phenomena that are not strictly considered to be "paranormal", and for being substantially culture-specific (with the content of these items often stemming from primarily Western legends/folklore). Including these items, as well as items reflecting potentially different belief clusters, may reduce the overall validity of scales attempting to assess paranormal beliefs, subsequently reducing the accuracy of these measures. While it is important to be clear about the latent construct of interest before construction of a psychometric assessment tool, ensuring appropriate and rigorous statistical assessment of scale properties during development will help to ensure that items effectively represent the same underlying dimension.

Ensuring psychometric assessment tools are reliable and valid is essential for accurate measurement of latent constructs in empirical research. The findings from Studies 3.1. and 3.2. highlight the importance of clear concept definition and rigorous statistical methods in the development of such tools. It is suggested that future research, both in psychology more widely and specifically related to paranormal beliefs, makes use of scales derived from modern test theory methods, e.g., the Paranormal and Supernatural Beliefs Scale (PSBS; Dean et al., 2021) put forward here as a new measure of belief in paranormal phenomena. This will help to improve the assessment of latent constructs, such as paranormal beliefs, so that experimental and correlational research can more accurately explore the relationships between these and other factors (e.g., cognitive function).

CHAPTER 4. PARANORMAL BELIEFS AND EXECUTIVE FUNCTIONS

INTRODUCTION

Executive functioning and Miyake's triad

Broadly defined, "*executive functioning*" refers to the higher order processes involved in the control and coordination of cognitions required for effective problem solving and decision-making in goal-oriented tasks. Executive functioning is a multidimensional construct, and three components are thought to be pertinent, including: set-shifting ability (referred to hereafter as "*cognitive flexibility*"), inhibitory control, and the updating and monitoring of working memory representations (Diamond, 2013; Miyake et al., 2000a).

Problem solving and decision making are goal-directed processes in which an individual will attempt to resolve a discrepancy between an initial state and a desired end state (Marsiske & Margrett, 2006). The types of problem situations an individual will encounter vary in their complexity and the cognitive effort required to resolve them. Flexibility of cognitive systems is therefore required to allow for effective problem solving and decision making. For such flexibility, an individual needs to have: an awareness that any given situation is likely to have multiple alternative solutions, a willingness to adapt to a given situation, and the belief that they can effectively adapt to a given situation (Martin & Anderson, 1998). *Cognitive flexibility* can be defined as the readiness of an individual to respond to relevant environmental stimuli, and to selectively use knowledge to adaptively fit decision-making needs (see Spiro et al., 1988; Scott, 1962). Put more simply, cognitive flexibility reflects an individual's ability to quickly change their way of thinking when shifting between tasks or problem situations. Consequently, the terms "*cognitive flexibility*" and "*set-shifting ability*" are often used interchangeably. In contrast, a lack of cognitive flexibility may represent an opposing *cognitive rigidity*, defined as the tendency of an individual to repeat the same thought or behaviour in response to a stimulus. This persistence, or perseveration, demonstrates an inability to shift between different concepts or to change behavioural responses. However, effective cognitive flexibility is thought to also rely on the other core executive functions of inhibitory control and working memory updating ability. For example, effectively shifting between different mental representations may require an individual to suppress (or inhibit) previous perspectives and provoke different perspectives (Diamond, 2013).

Inhibitory control aids successful completion of goal-directed tasks through the suppression of goal-irrelevant responses and stimuli (Tiego et al., 2018). This can be described as *response inhibition* (concerning the suppression of behavioural or emotional responses to a stimulus that would interfere with task performance), and *attentional inhibition* (involving the suppression of attention to irrelevant stimuli or thoughts unrelated to the task; Tiego et al., 2018; Diamond, 2013). Inhibitory control is also an important cognitive component of successful social interaction, allowing for the suppression of socially inappropriate or insensitive responses (von Hippel & Gonsalkorale, 2005). Failures of inhibitory control are often seen in children and the elderly, as well as in a range of clinical and behavioural disorders (Munakata et al., 2011). These failures may be observed as impulsive behaviours, conditioned responses (i.e., "old habits"), and a lack of focused attention (Diamond, 2013).

In contrast to the suppression of information, the ability to effectively monitor and update working memory representations requires information to be appropriately coded in relation to its relevance to the task, with this information held in working memory representations and revised or replaced with more relevant information as per task demands (Miyake et al., 2000A; Morris & Jones, 1990). In this sense, *working memory updating ability* requires not only the passive maintenance of task-relevant information, but the active manipulation of this information (Miyake et al., 2000A). Poor working memory updating ability shows some similarity with ineffective cognitive flexibility, manifesting as rumination and perseveration (Kaiser et al., 2015; Amso et al., 2014; Meiran et al., 2011).

Models of executive functioning

Despite cognitive flexibility, inhibitory control, and working memory updating ability being identified as core executive functions, the separability of these functions is still debated. This raises conceptual limitations related to the precision of definitions and inconsistent use of terminology (whereby different research groups may use the same term to refer to conceptually different executive functions or use different terms to refer to the same EF; Miyake et al., 2000B). While some argue that these executive functions emerge from a single, larger psychological construct such as general or fluid intelligence, reasoning ability, or processing speed (see Banich, 2009), correlational studies suggest some partial separability of executive functions. Despite participant scores on different executive tests often correlating positively, the correlations are typically low-to-moderate in strength (*r*s between .4 to .6; Gilbert & Burgess, 2008; Miyake et al., 2000b), suggesting partially distinct functions that share common processes. Neurocognitive studies also provide support for *some* separability of executive functioning, although the neural mechanisms underpinning executive functions are far from clear. General agreement exists that the frontal lobes play a key role in executive functions, with a particular emphasis on the prefrontal cortex (PFC).

The PFC is a heterogenous neocortical region sharing connections with sensory and motor systems, as well as a range of subcortical structures (e.g., the thalamus), and plays a central role in the integration of information needed for complex behaviour (Miller & Cohen, 2001; Miller, 2000). The PFC is thought to be particularly important for the cognitive control required in goal-oriented tasks, although the exact nature of its involvement with executive functions is unclear. Some posit that distinct regions of the PFC are related to specific executive functions. For example, Brodmann Areas (BA) 18 and 19 (relating to the occipital cortex) have been related to cognitive flexibility (Panikratova et al., 2020; Kim et al., 2011), while BA 6 and 8 (relating to the premotor cortex and frontal eye fields, respectively) have been associated with inhibitory control (Rajesh et al., 2021). Others, however, suggest that PFC regions adapt to task demands rather than carry out predetermined functions (Duncan & Owen, 2002). It has also been argued that individual functions may be assigned to subregions of the PFC but are coordinated by an overarching PFC function that is not localised to a distinct subregion and processes temporally complex events (such as when information crucial to task performance or learning can only be interpreted with respect to a preceding event; Wilson et al., 2010).

While some disagreement exists surrounding models of executive functioning, it has been suggested that combining neurobiological, psychological, and computational models will produce a more integrated account that will not only aid understanding of these cognitive processes, but provide more fruitful avenues for interventions related to executive deficit (Banich, 2009).

Issues relating to the measurement of executive functions

In addition to conceptual and theoretical limitations, several issues arise in relation to the measurement of executive functions. Of particular note is the so-called "task impurity problem". As executive functioning describes a variety of cognitive abilities required for effective problem-solving and goal-directed behaviours, tasks designed to assess executive functions inevitably draw on non-executive abilities such as visuospatial processing (Miyake et al., 2000b). A low score on one executive test may not therefore necessarily reflect a 'pure' executive deficit. The Wisconsin Card Sorting Test (WCST; see Berg, 1948; Grant & Berg, 1948), one of the most frequently used executive tests, demonstrates the task impurity problem well. The WCST requires participants to sort a deck of abstract cards according to an unspecified categorisation rule (e.g., colour). The rule by which participants must categorise the cards changes after a given number of trials or correct responses, requiring participants to then decipher the new rule. While this task requires effective cognitive flexibility to shift between different rules, successful completion of the task also relies on some inhibitory control of behavioural responses (particularly during a rule change), as well as working memory updating ability to revise and replace the current categorisation rule held in working memory representations (see Roca et al., 2010; Miyake et al., 2000b). The task also requires effective visual, numerical, and lexical processing to determine categorisation rules and respond to feedback relating to task performance. A deficit relating to any of these abilities would confound scores on the test and produce scores that appear indicative of an executive deficit.

A further issue relating to the measurement of executive functions lies in the heterogeneity of assessment tools. In their systematic review of 106 studies, Baggetta and Alexander (2016) found 11 different test batteries (performance-based tasks and behavioural rating scales) and 109 different individual behavioural tasks used to assess executive functioning. Of the 109 behavioural tasks, 56 were only utilised once, suggesting that researchers often develop novel measures of executive functioning. For these novel measures to be deemed reliable, effective pilot testing and statistical evaluation of their psychometric properties needs to be undertaken prior to their use. The lack of standardised testing for executive functions also hinders cross-study comparisons, and often results in researchers using the same tasks to assess different aspects of executive functioning. For example, the Stroop task (Stroop, 1935) has been used as a measure of inhibition, working memory, cognitive control, attentional control, and general executive functioning (Baggetta & Alexander, 2016). Not only does this again impact on the ability to compare findings across studies, but further highlights the task impurity problem with executive tasks being used to tap into various cognitive functions.

Executive functions and paranormal beliefs

As discussed, the study of executive functions is far from simple; however, some evidence exists of a relationship between paranormal beliefs and executive difficulties. Wain and Spinella (2007) report a significant negative relationship between paranormal beliefs and impulse (inhibitory) control using a self-report measure of executive functions, while Lindeman et al. (2011) report significantly more errors on the WCST for believers compared to sceptics (suggesting an association between paranormal beliefs and reduced cognitive flexibility). Despite the small volume of literature explicitly assessing executive functions in relation to paranormal beliefs, many of the cognitive function studies conducted in this area could be considered as indirectly assessing executive functioning (as noted in the Systematic Review chapter of this thesis). This may be related to both the task impurity problem previously discussed here, and the fluid-executive theory (Dean et al., 2022; see the Systematic Review chapter of this thesis).

The present studies

The present studies build upon the small foundation of work explicitly exploring executive deficits relating to paranormal beliefs.

The first study explores Miyake et al.'s (2000) triad of executive functions (cognitive flexibility, inhibitory control, and working memory updating ability). Owing to the small volume of literature and a lack of theory surrounding cognitive functions and paranormal beliefs, this study did not make specific hypotheses relating to the direction of the relationship between executive functioning and paranormal beliefs. As schizotypal thinking has also been shown to have strong links with both paranormal beliefs (see the General Introduction chapter of this thesis) and with executive functioning (see Steffens et al., 2018, for more detailed discussions), the study also aimed to provide further support for these relationships. It was hypothesised that high schizotypy scores would be associated with higher paranormal beliefs, and poorer executive functioning.

The second study aims to replicate Lindeman et al.'s (2011) cognitive flexibility study in an appropriately powered sample. This study used two online behavioural measures of set-shifting ability (including an adapted WCST) in combination with a self-report measure of cognitive flexibility. It was anticipated that high paranormal belief scores would correlate with inhibited performance on both behavioural measures of cognitive flexibility, although no directional hypotheses were made in relation to the self-report measure of cognitive flexibility.

The final studies in this chapter concern replications and extensions of Study 4.2., attempting to further clarify the relationship between paranormal beliefs and cognitive flexibility.

STUDY 4.1. PARANORMAL BELIEFS, EXECUTIVE FUNCTIONS AND DELUSIONAL IDEATION IN HEALTHY INDIVIDUALS

METHOD

Participants

An opportunistic sample of 40 participants were recruited through advertisements placed on social media and the University of Hertfordshire's online study recruitment website. Advertisements asked for participants over the age of 18 and fluent in English to complete three multiple-choice questionnaires (about their belief in paranormal phenomena, vivid mental experiences and cognitive difficulties), in addition to three cognitive tasks. The sample consisted of 31 females (78%), 8 males (20%) and one unreported, aged between 18-50 (M = 22.95, SD = 7.48). Most participants were white (48% white British, 12.50% other white background, 03% white Irish), and had a post-secondary education (60%). Of the participants with a university education (15% undergraduate degree, 18% postgraduate degree), most came from a psychology discipline (90%). See Table 4.1. for full demographic details.

Materials

Questionnaires

The Revised Paranormal Belief Scale (RPBS; Tobacyk, 2004) was used to assess participants' belief in paranormal phenomena. The scale consists of 26 items (e.g., "there are actual cases of witchcraft") and contains seven subscales designed to capture belief in a range of phenomena: Traditional Religious Belief, Psi, Witchcraft, Superstition, Spiritualism, Extraordinary Life Forms and Precognition. Participants rate the extent to which they agree with each statement on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The RPBS demonstrates an excellent test-retest reliability over a four-week period (r = .92), and strong test-retest reliability coefficients shown for each subscale: Traditional Religious Belief .95, Psi .71, Witchcraft .93, Superstition .89, Spiritualism .91, Extraordinary Life Forms .91, and Precognition .81 (Tobacyk, 2004).

Peters et al. Delusions Inventory (PDI; Peters et al., 2004, see Appendix M) consists of 21 items and was designed to measure delusional ideation in the general population. The PDI yields four separate scores: a yes/no score, a distress score, a preoccupation score, and a conviction score. Participants respond to each item (e.g., "do you ever feel that people look at you oddly because of your appearance?") with a 'yes' (1) or 'no' (0), giving a maximum possible yes/no score of 21. Distress, preoccupation and conviction are rated on 5-point Likert scales, and a 'no' response to the original item yields an automatic score of 0 for these items. The maximum possible score for each dimension is 105. A grand total PDI score can also be obtained by adding the yes/no scores and the scores from each dimension, to yield a maximum possible score of 336. The whole scale has very good internal reliability, with a

Cronbach's alpha of .82, and shows good test-retest reliabilities for each subscale: yes/no .78, distress .81, preoccupation .81 and conviction .78 (Peters et al., 2004).

The Dysexecutive Questionnaire (DEX; Burgess et al., 1996, see Appendix N) was used to measure participants' perception of their own executive functioning and dysexecutive problems. The DEX contains 20 items (e.g., "I have difficulty thinking ahead or planning for the future") which participants respond to using a 4-point Likert scale (1 = never, 4 = very often). The scale contains three subscales (Behaviour, Cognition and Emotion) and has a maximum possible score of 80. The DEX demonstrated very good internal reliability in this sample, with a Cronbach's alpha of .86.

Behavioural tests

Participants were first presented with a backward digit span (BDS) test to assess their updating ability, during which they are verbally presented with a string of random digits. The length of each string starts at two digits, and gradually increase by one until participants consecutively fail two trials of the same length. Participants are required to repeat the digits aloud, recalling the strings in reverse order. For example, if participants are given the string "5, 7, 4, 6", the correct response would be "6, 4, 7, 5". The BDS test is a commonly used and recognised measure of verbal working memory updating, and previous studies have demonstrated good reliability of this task through Cronbach's alpha (.69; Engle et al., 1999) and correlation coefficients (.78; Henry, 2001).

The Hayling Sentence Completion Test (HSCT; Burgess & Shallice, 1997) was used to assess participants' inhibitory control. The HSCT requires participants to complete sentences with a semantically connected word in Part A, and a semantically unconnected word in Part B. Therefore, in Part B participants must inhibit a natural response to the item and instead generate a random response. For example, in Part A, a correct response to the item "the rich child attended a private _____" might be the word "school". In Part B, a correct response to the item "London is a very busy _____" would require participants to inhibit the words "city" or "town" and instead produce a semantically unconnected word (e.g., "teapot"). Performance on the HSCT is measured through participants' speed in Part A and Part B, and the number of errors made in Part B. Participants' errors in Part B are coded as either a Category A error (providing a semantically connected word to complete the sentence, such as "city" for the example above), or a Category B error (a word which is somewhat semantically connected to the sentence, such as "village" for the example above). Speed and accuracy are used to produce scaled scores, which when added together (speed in Part A + speed in Part B + accuracy in Part B) give an overall total score for the HSCT. The overall score is then scaled from 1-10 (1 = impaired, 10 = very superior), with higher overall scores producing a higher scaled score.

Finally, the Brixton Spatial Anticipation Test (BSAT; Burgess & Shallice, 1997) was used to assess participants' cognitive flexibility across 55 trials. Participants are presented with two rows of five circles, labelled from 1-10, and are asked to predict the location the coloured circle will be on the next page. The rule influencing the movement of the target circle changes randomly, and participants are required to detect and apply the current rule. As with the HSCT, participants' total number of errors are used to produce a scaled score from 1-10 (1 = impaired, 10 = very superior), with the fewest errors producing a higher scaled score. While there is little statistical information published on the internal consistency and reliability of the BSAT and HSCT measures (two widely used dysexecutive measures), they have been reported to have moderate Cronbach's alpha values in patient samples (.70; Crawford & Henry, 2005).

Procedure

The three questionnaires and demographic items were administered as an online survey using Qualtrics Survey Software (Qualtrics, Provo, UT; https://www.qualtrics.com).

Participants were informed that the study aimed to explore relationships between executive functioning, paranormal beliefs, and vivid mental experiences. Respondents were asked to provide basic demographic information, including their age, gender, ethnicity, level of education and academic discipline if they had indicated a university education (see Table 2.1. for full demographic response options). All participants then completed the questionnaires in the same order (RPBS, PDI and the DEX), before completing the three executive tests in the same order (BDS, HSCT, and the BSAT).

Informed consent was obtained from all participants and ethical approval for the study was granted by the University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (HSET ECDA; protocol number LMS/PGR/UH/03844).

Statistical analyses

A power analysis conducted using G*Power 3.1.9.4. (Faul, et al., 2007) indicated that a minimum of 82 participants would be needed to detect a medium effect with 80% power, at an alpha of .05. However, face-to-face testing was halted owing to Covid-19 restrictions, resulting in a smaller sample of 40 participants. To account for the potential underpowering of the study, both frequentist (*p* values) and Bayesian (Bayes factor) methods of inference were used. Bayesian methods have been suggested to be suitable for samples of all sizes (Wagenmakers et al., 2018b). The Bayes factor (BF_{10}) assesses evidence for the alternative hypothesis (H_1) compared to the null hypothesis (H_0) given the observed data (Wagenmakers et al., 2018a). For example, a BF_{10} of 40 indicates that the observed data are 40 times more likely under H_1 than under H_0 and provides very strong evidence for H_1. However, a BF_{10} of 0.40 indicates that the observed data are 40 times more likely under H_0 than H_1 and provides anecdotal evidence for H_0. Owing to the underpowered sample and the strength of Bayesian statistics for such data, Bayes factors will be used to determine the strength of any significant results found through frequentist analysis.

RESULTS

Table 4.1. presents a summary of participant demographic information. Of the sample, most were white, well-educated females with a background in psychology.

Table 4.1. Frequencies and percentages of participant demographics.

Variable	n	%
Gender		
Male	8	20.0
Female	31	77.5
Prefer not to say	1	2.5
Ethnicity		
African	1	2.5
Asian/Asian British	2	5.0
Bangladeshi	1	2.5
Chinese	2	5.0
Indian	1	2.5
Pakistani	6	15.0
White British	19	47.5
White Irish	1	2.5
Mixed background	2	5.0
Other White background	5	12.5
Education		
Vocational	1	2.5
Secondary education	2	5.0
Post-secondary education	24	60.0
Undergraduate degree	6	15.0
Postgraduate degree	7	17.5
Discipline		
Art & Humanities	1	2.5
Education	1	2.5
Psychology	36	90.0
Other	1	2.5
N/A (have not attended university)	1	2.5

Descriptive statistics

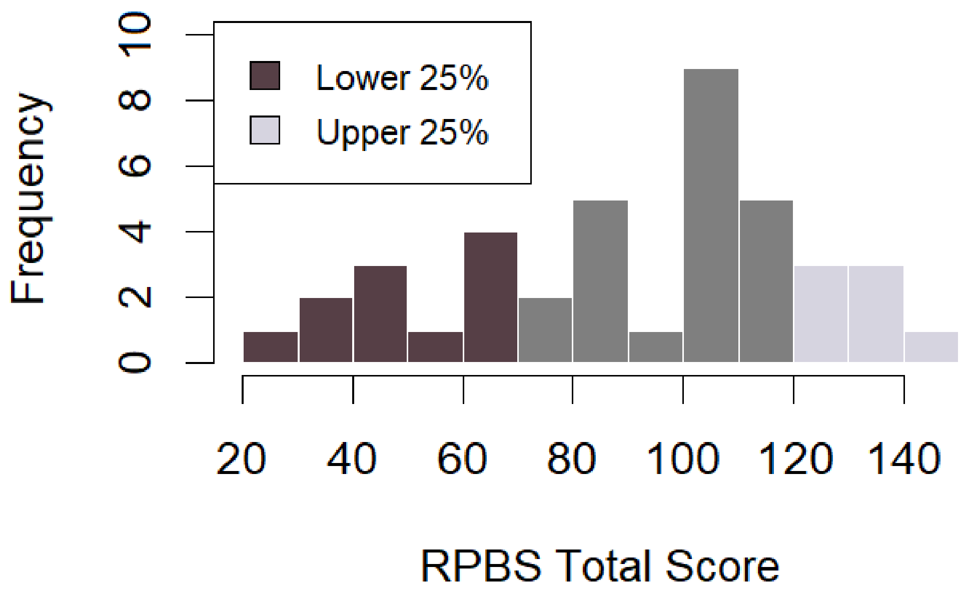
Means and standard deviations, as well as reliability coefficients (Cronbach's alpha), were calculated for all three scales. Cronbach's alpha was .93 for the RPBS, .94 for the PDI, and .86 for the DEX. The descriptive statistics for all measures, including the executive tests, can be found in Table 4.2. To visually represent the distribution of RPBS scores, a histogram was created in RStudio (RStudio Team, 2020), and the results can be found in Figure 4.1.

Table 4.2. Cronbach's alpha coefficients, Means (SD), and scale ranges for the questionnaires and executive tests.

Variable	*α*	*M*	*SD*	Scale range
1. RPBS	.93	92.58	30.34	26-182
2. PDI	.94	72.20	43.65	0-336
3. DEX	.86	30.28	11.82	20-80
4. HSCT	-	5.70	1.34	1-10
5. BSAT	-	6.75	2.11	1-10
6. BDS	-	6.55	2.39	-

Note: RPBS = Revised Paranormal Beliefs Scale, PDI = Peters et al. Delusions Inventory, DEX = Dysexecutive Questionnaire, HSCT = Hayling Sentence Completion Test, BSAT = Brixton Spatial Anticipation Test, BDS = Backwards Digit Span.

Figure 4.1. Frequency of total RPBS scores.



Note: RPBS = Revised Paranormal Belief Scale

Predicting paranormal belief

Frequentist regression analyses

Multiple linear regressions were conducted with RPBS scores at the outcome measure. Total PDI and DEX scores, as well as participants' scores on the three executive function tests (BDS, Hayling and Brixton), were used as the five predictor variables. The overall model was statistically significant ($F(5, 34) = 7.518, p < .001; R^2 = .53$). PDI scores were a significant, positive predictor of RPBS scores, while BSAT scores were identified as a significant negative predictor of RPBS scores. However, the DEX, BDS and HSCT were not significant predictors of RPBS scores. Table 4.3 presents a summary of the multiple regression statistics.

Table 4.3. Summary of multiple linear regression statistics for paranormal belief.

	B	SE	*β*	*t*	*p*	*R²*	Adj. *R²*
Model						**.525**	**.455**
PDI	.32	.12	.46	2.75	.009**		
DEX	-.15	.39	-.06	-.38	.704		
BDS	-2.85	1.62	-.22	-1.77	.086		
HSCT	-.50	2.97	-.02	-.17	.868		
BSAT	-4.91	2.01	-.34	-2.45	.020*		

*Note: PDI = Peters et al. Delusions Inventory, DEX = Dysexecutive Questionnaire, BDS = Backwards Digit Span, HSCT = Hayling Sentence Completion Test, BSAT = Brixton Spatial Anticipation Test, *** = $p \leq .001$, ** = $p \leq .01$, * = $p \leq .05$*

To find a reduced model that best explains the data, a backward stepwise regression analysis was conducted. After the removal of non-significant results (HSCT, DEX and BDS, respectively), the final model included two predictors and was statistically significant ($F(2, 37) = 16.347, p < .001; R^2 = .469$). The PDI significantly, positively predicted RPBS scores, and BSAT scores significantly, negatively predicted RPBS scores. The summary of statistics from each multiple regression are presented in Table 4.4 and Table 4.5, with the final model presented in Table 4.6.

Table 4.4. Summary of the multiple linear regression statistics for paranormal belief with HSCT scores removed.

	B	SE	β	t	p	R^2	Adj. R^2
Model						**.525**	**.470**
PDI	.32	.11	.45	2.79	.008**		
DEX	-.15	.39	-.06	-.39	.696		
BDS	-2.93	1.54	-.23	-1.90	.065		
BSAT	-5.02	1.86	-.35	-2.70	.011*		

*Note: PDI = Peters et al. Delusions Inventory, DEX = Dysexecutive Questionnaire, BDS = Backwards Digit Span, HSCT = Hayling Sentence Completion Test, BSAT = Brixton Spatial Anticipation Test, *** = p ≤ .001, ** = p ≤ .01, * = p ≤ .05*

Table 4.5. Summary of the multiple linear regression statistics for paranormal belief with HSCT and DEX scores removed.

	B	SE	β	t	p	R^2	Adj. R^2
Model						**.523**	**.483**
PDI	.29	.09	.41	3.32	.002**		
BDS	-3.02	1.50	-.24	-2.01	.052		
BSAT	-5.23	1.76	-.36	-2.97	.005**		

*Note: PDI = Peters et al. Delusions Inventory, DEX = Dysexecutive Questionnaire, BDS = Backwards Digit Span, HSCT = Hayling Sentence Completion Test, BSAT = Brixton Spatial Anticipation Test, *** = p ≤ .001, ** = p ≤ .01, * = p ≤ .05*

Table 4.6. Summary of the final multiple linear regression statistics for paranormal belief following backward stepwise analysis.

	B	SE	β	t	p	R^2	Adj. R^2
Model						**.469**	**.440**
PDI	.32	.09	.46	3.63	.001***		
BSAT	-5.38	1.83	-.37	-2.94	.006**		

*Note: PDI = Peters et al. Delusions Inventory, BSAT = Brixton Spatial Anticipation Test *** = p ≤ .001, ** = p ≤ .01, * = p ≤ .05*

Bayesian regression analyses

Bayesian multiple linear regression analyses were conducted to further analyse the data, due to the small sample size and multiple non-significant findings within the original frequentist model. BF_{10} showed that the data were 9.396 times more likely under a model containing the PDI, BDS and BSAT than under the null model (H_0), indicating moderate evidence (see Wagenmakers et al., 2018a). The BF_{10} for the next strongest model (PDI and BSAT only) showed that the data were 6.393 more likely under this model than under H_0, indicating moderate evidence. All other models identified through the Bayesian analysis indicated either anecdotal evidence for the alternative hypothesis (H_1) or anecdotal evidence for H_0.

Correlational analyses

Frequentist correlations

Pearson product-moment correlation revealed a positive correlation between RPBS scores and PDI scores ($r(38) = .59, p < .001$), and a negative correlation between RPBS scores and both BSAT scores ($r(38) = -.53, p < .001$) and BDS scores ($r(38) = -.37, p = .020$). No significant correlation emerged between RPBS scores and DEX scores or HSCT scores. A significant positive correlation was also seen between the PDI and the DEX ($r(38) = .57, p < .001$), as well as a significant negative correlation between PDI and BSAT scores ($r(38) = -.34, p = .034$). The full correlation matrix is presented in Table 4.7.

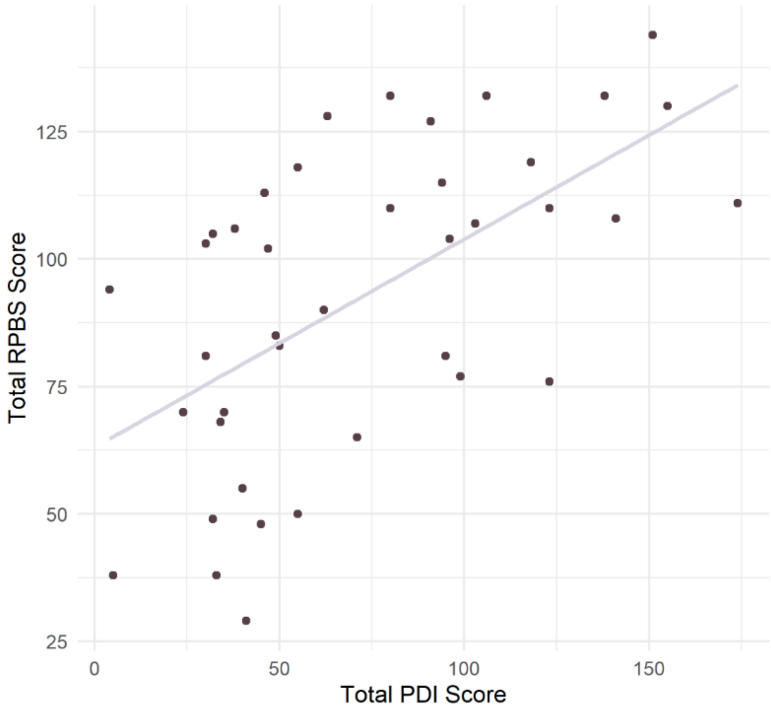
Table 4.7. Pearson product-moment correlations between RPBS, PDI, DEX, BDS, HSCT and BSAT scores.

Variable	1	2	3	4	5	6
1. RPBS		.59**	.19	-.37*	-.20	-.53**
2. PDI			.55**	-.22	.00	-.34*
3. DEX				-.00	.14	.03
4. BDS					.26	.11
5. HSCT						.34*
6. BSAT						

*Note: RPBS = Revised Paranormal Beliefs Scale, PDI = Peters et al. Delusions Inventory, DEX = Dysexecutive Questionnaire, BDS = Backwards Digit Span, HSCT = Hayling Sentence Completion Test, BSAT = Brixton Spatial Anticipation Test, ** = p < .001, * = p < .05*

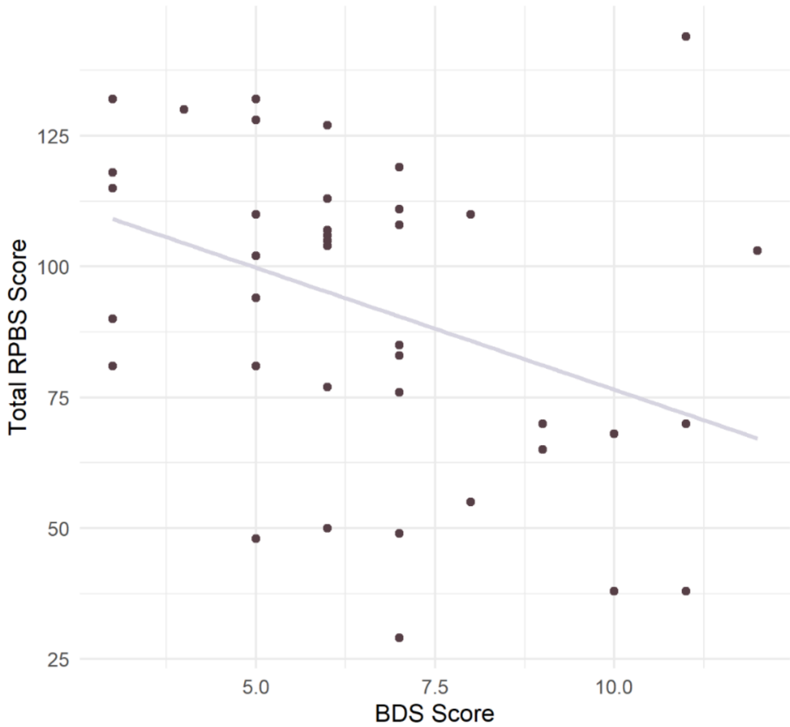
To visually represent the significant Pearson product-moment correlations, scatterplots were made in RStudio (RStudio Team, 2020), and the results can be found in Figure 4.2 – Figure 4.7.

Figure 4.2. Pearson product-moment correlation between RPBS and PDI scores.



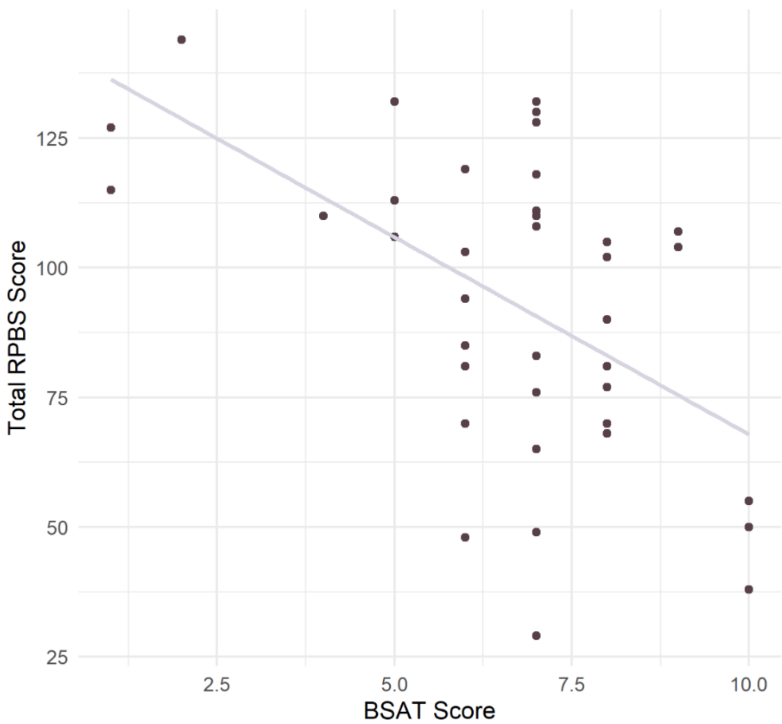
Note: RPBS = Revised Paranormal Beliefs Scale, PDI = Peters et al. Delusions Inventory.

Figure 4.3. Pearson product-moment correlation between RPBS and BDS scores.



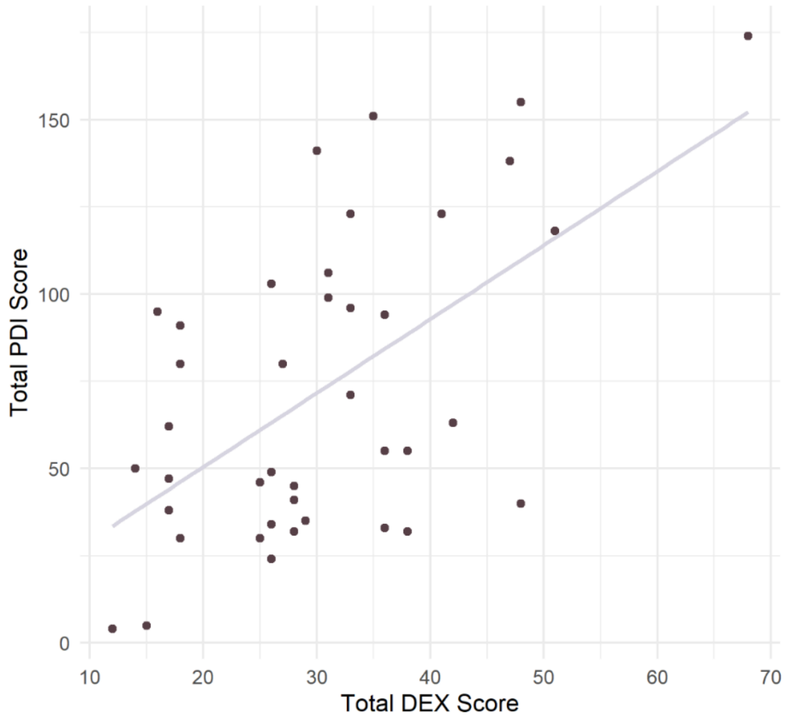
Note: RPBS = Revised Paranormal Beliefs Scale, BDS = Backwards Digit Span.

Figure 4.4. Pearson product-moment correlation between RPBS and BSAT scores.



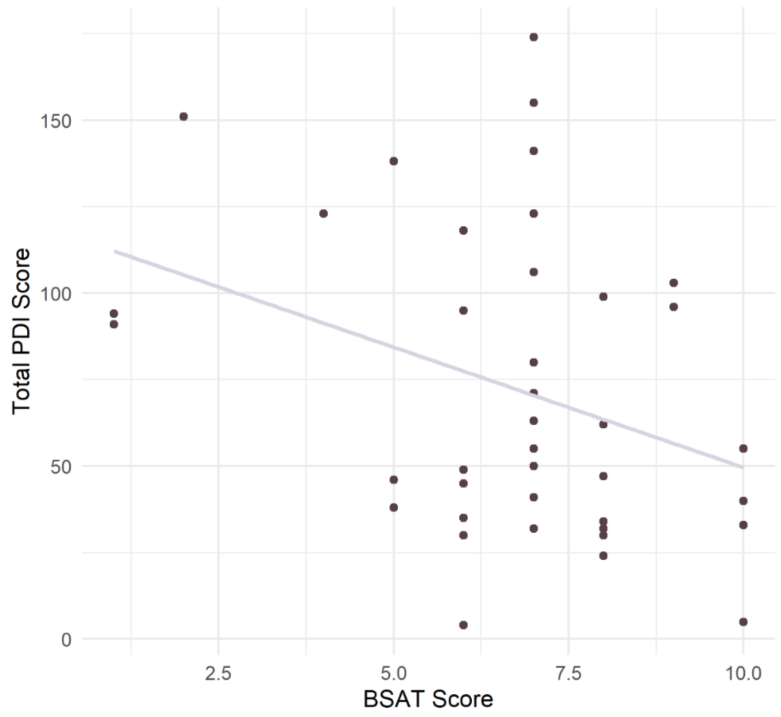
Note: RPBS = Revised Paranormal Beliefs Scale, BSAT = Brixton Spatial Anticipation Test.

Figure 4.5. Pearson product-moment correlation between PDI and DEX scores.



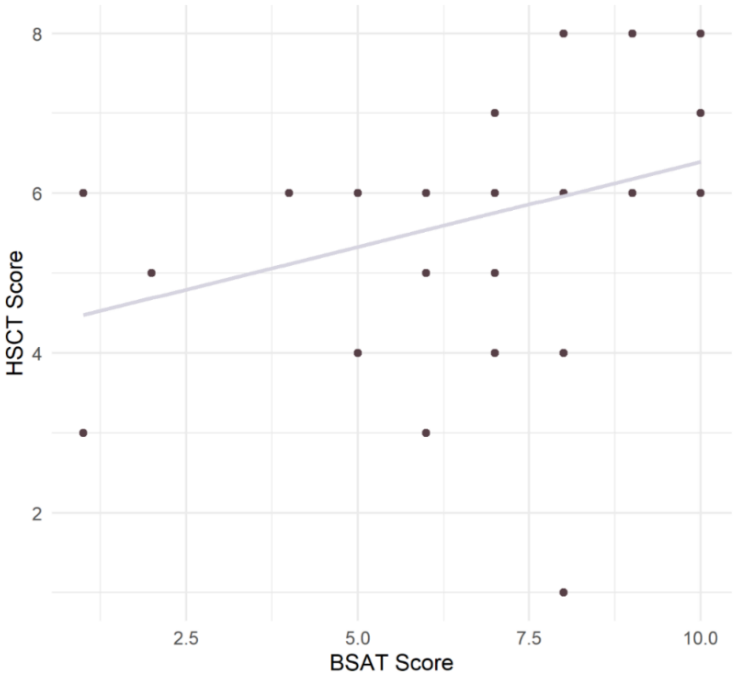
Note: PDI = Peters et al. Delusions Inventory, DEX = Dysexecutive Questionnaire.

Figure 4.6. Pearson product-moment correlation between PDI and BSAT scores.



Note: PDI = Peters et al. Delusions Inventory, BSAT = Brixton Spatial Anticipation Test.

Figure 4.7. Pearson product-moment correlation between HSCT and BSAT scores.



Note: HSCT = Hayling Sentence Completion Test, BSAT= Brixton Spatial Anticipation Test.

Bayesian correlations

Examination of the Bayesian Pearson product-moment correlation matrix revealed a significant positive correlation between RPBS and PDI scores ($r(38) = .59$, $BF_{10} = 418.187$). The Bayes factor for the correlation between RPBS and PDI scores is indicative of extreme evidence for the alternative hypothesis (H_1). A negative correlation was found between RPBS and BSAT scores ($r(38) = -.53$, $BF_{10} = 74.764$), with the Bayes factor indicating very strong evidence for H_1. Finally, a positive correlation was found between the PDI and the DEX ($r(38) = .57$, $BF_{10} = 270.434$), with the Bayes factor indicating extreme evidence for H_1. The Bayes factor for the correlation between RPBS scores and BDS scores ($r(38) = -.37$, $BF_{10} = 2.683$) indicated anecdotal evidence for H_1. Finally, the correlations between PDI and HSCT scores, BSAT and BDS scores, and the correlations between DEX scores and all three executive tests (BDS, HSCT and BSAT) demonstrated Bayes factors between .197 and .283, indicating moderate evidence for H_0. All other correlations in the Bayesian Pearson product-moment correlation matrix demonstrated Bayes factors between 0.385 and 1.726, indicating either anecdotal evidence for H_0 or no evidence for either H_0 or H_1. The full correlation matrix is presented in Table 4.8.

Table 4.8. Bayesian correlations between RPBS, PDI, DEX, BDS, HSCT and BSAT Scores (with Pearson's *r* values and Bayes factors).

Variable		1	2	3	4	5	6
1. RPBS	*r*		.59***	.19	-.37	-.20	-.53**
	BF_{10}		418.19	.39	2.68	.42	74.76
2. PDI	*r*			.57***	-.22	.00	-.34
	BF_{10}			270.43	.47	.20	1.73
3. DEX	*r*				-.00	.14	.03
	BF_{10}				.20	.28	.20
4. BDS	*r*					.26	.11
	BF_{10}					.71	.25
5. HSCT	*r*						.34
	BF_{10}						1.69
6. BSAT	*r*						
	BF_{10}						

*Note: RPBS = Revised Paranormal Beliefs Scale, PDI = Peters et al. Delusions Inventory, DEX = Dysexecutive Questionnaire, HSCT = Hayling Sentence Completion Test, BSAT = Brixton Spatial Anticipation Test, BDS = Backwards Digit Span, *** = $BF_{10} > 100$, ** = $BF_{10} > 30$, * = $BF_{10} > 10$*

DISCUSSION OF STUDY 4.1.

Despite the small sample in Study 4.1., some support was found for a negative relationship between paranormal beliefs and executive functioning. Cognitive flexibility was negatively related to paranormal beliefs, with this finding supported by both Bayesian and frequentist statistics. Specifically, Bayesian analyses suggested very strong evidence for a negative correlation between cognitive flexibility and paranormal beliefs. A negative relationship between paranormal beliefs and updating ability was also suggested by frequentist analyses, however Bayesian analyses indicated only anecdotal evidence for this relationship. Neither Bayesian nor frequentist analyses identified any significant relationships for paranormal beliefs and inhibitory control. Paranormal beliefs were significantly associated with increased schizotypal thinking, while schizotypal thinking was associated with poorer working memory updating ability and self-rated cognitive flexibility scores. Frequentist analyses suggested a significant negative relationship between schizotypal thinking and cognitive flexibility (such that higher schizotypal thinking was associated with reduced cognitive flexibility), but this was not found through subsequent Bayesian assessment. Neither the frequentist nor the Bayesian analyses supported a relationship between schizotypal thinking and reduced inhibitory control. While these results were unexpected, they do provide some support for previous research suggesting that individuals with high levels of schizotypal thinking rate their executive functioning as worse than those with lower levels of schizotypal thinking, but do not always demonstrate such deficits on behavioural tests of executive functioning (Laws et al., 2008).

While the sample size for this study poses potential limitations, Bayesian statistics were able to provide some additional assessment of the findings highlighted from the frequentist analyses. These findings demonstrate some consistency with previous research suggesting that paranormal beliefs are associated with executive difficulties. In particular, the findings are comparable to those from Lindeman et al. (2011), which demonstrated reduced cognitive flexibility in paranormal believers compared to sceptics. The present findings do, however, also demonstrate some inconsistencies with previous research in this area. Work by Wain and Spinella (2007) suggested reduced inhibitory control associated with paranormal beliefs; however, no support for this association was found in the present study. It is possible that the present work failed to replicate this finding due to differences executive assessment, with Wain and Spinella's (2007) study using a self-report measure and the present work using behavioural measures. It is possible, therefore, that paranormal believers rate their inhibitory control as worse, despite demonstrating normal performance on inhibitory tasks. Paranormal believers, therefore, may demonstrate an impairment of executive awareness relating to inhibition, rather than a behavioural deficit of the function. However, self-rated executive difficulties, as measured by the DEX, were not related to paranormal beliefs in the present study.

Interpretation of the present findings should also consider the limitations of the selected executive tests. For example, while the weak support for a relationship between paranormal beliefs and working memory updating ability might be attributed to the small sample size, previous work has suggested that backwards digit span salience is significantly affected by external factors. For example, as the backwards digit span task is administered verbally by the researcher, variations in digit enunciation (such as small changes in the pitch or rhythm of the presentations) have been shown to influence participants' performance (see Silverman, 2007). One of the main limitations of both the Hayling and Brixton tasks is the use of Sten (Standard Ten) scoring. These scores are a commonly used metric in psychometric testing and do provide simplistic outcomes that are interpretable for lay audiences. The difference between Sten scores, however, corresponds to around .50 of a standard deviation, and may therefore obscure meaningful differences that might be present in participants' raw scores (Crawford & Henry, 2005).

While the present findings suggest a potential relationship between paranormal beliefs and working memory updating ability, the strongest evidence was found for cognitive flexibility. Subsequent studies were therefore considered to further assess this relationship with more sensitive executive measures and larger samples.

STUDY 4.2. PARANORMAL BELIEFS AND COGNITIVE FLEXIBILITY

The aim of Study 4.2. was to replicate and extend the findings from Study 4.1. using more sensitive measures of set-shifting ability and a larger sample size. In particular, the study aimed to find a relationship between paranormal beliefs and reduced cognitive flexibility; and to further assess the relationship between schizotypal thinking and reduced cognitive flexibility. As Study 4.1. provided some support with the DEX for previous findings suggesting that individuals with high levels of schizotypal thinking rate their executive functioning as worse than those with lower levels of schizotypal thinking, but do not always demonstrate deficits on behavioural tests (Laws et al., 2008), the present study also included a self-report measure of cognitive flexibility to see if this finding could be replicated for a specific executive domain. A-priori power analysis indicated that a minimum of 98 participants were required to detect a medium effect with 80% power at an alpha of .05, while 123 participants would be needed to detect a medium effect with 90% power at an alpha of .05.

METHOD

Participants

Data were collected online using a sample of 123 participants consisting of undergraduate students from the University of Hertfordshire and members of the public. Prior to analysis, two participants were removed as they did not meet the study's eligibility criteria (ages were < 18). An additional participant was removed because of their outlying reaction time data for the overall time taken to complete the Wisconsin Card Sorting Test ($>$ 4 standard deviations above the mean). The final sample therefore consisted of 120 participants (12 males, 106 females, 2 unreported, aged between 18 and 60; M = 23.90, SD = 8.79). Of the final sample, the majority (48%) had a post-secondary education, followed by an undergraduate education (23%), secondary education (22%), postgraduate education (04%), and vocational education (01%), with three participants (02%) choosing not to provide this demographic information.

Materials

Questionnaires

Paranormal beliefs were assessed through a 14-item version of the Paranormal and Superstitious Beliefs Scale (PSBS-14; see Dean et al., 2021). Participants rated their level of agreement with each item (e.g., "it is possible to become possessed by an evil supernatural entity") using a 7-point Likert scale. Total scores for the PSBS-14 could range from 14-98, with the scale demonstrating strong internal consistency (α = .95; Dean et al., 2021).

Schizotypal thinking was measured using the Schizotypal Personality Questionnaire Brief (SPQ-B; Raine & Benishay, 1995, see Appendix O), adapted for use with a 5-point Likert response format (see Cohen et al., 2010; Wuthrich & Bates, 2005). Participants rated their level of agreement with items such as "some people find me a bit vague and elusive during a conversation". The 22-item scale provides a total score and individual scores for three subscales: cognitive-perceptual deficits (8 items), interpersonal deficits (8 items), and disorganisation (6 items). The Likert version of the SPQ-B has demonstrated good internal consistency for the total scale (α = .90), and for each of the subscales (cognitive-perceptual α = .79, interpersonal deficits α = .86, and disorganisation α = .83; Cohen et al., 2010).

The 20-item Cognitive Flexibility Inventory (CFI; Dennis & Vander Wal, 2010, see Appendix P) was completed by participants to self-assess cognitive flexibility. Participants rated their level of agreement with items (e.g., "I have a hard time making decisions when faced with difficult situations") using a 7-point Likert scale. As well as providing a total score, the CFI provides scores on two subscales: alternatives (13 items measuring the ability to perceive multiple alternative explanations for experiences and behaviours, and generate multiple alternative solutions to scenarios), and control (seven items measuring the tendency to perceive difficult situations as controllable). The CFI has presented with good internal consistency for the total scale (α = .91) and for the two subscales (alternatives α = .91, and control α = .84; Dennis & Vander Wal, 2010). The scale also demonstrates good convergent construct validity with other older measures of cognitive flexibility (e.g., the Cognitive Flexibility Scale; Martin & Rubin, 1995).

Behavioural tests

Due to the previously outlined limitation of the Brixton's Sten scores, and its limited capacity to be adapted for online use (necessary for compliance with Covid-19 restrictions in force at the time of data collection), cognitive flexibility in Study 4.2. was measured with two alternative tasks: a cued task-switching task, and an adapted version of the widely used Wisconsin Card Sorting Task.

The cued task-switching task (CTST; see Gollan et al., 2014; Capa et al., 2013, for examples of cued task-switching tasks) presented coloured shapes centrally on the screen as target stimuli against a white background. Participants were required to perform one of two possible tasks in each trial (referred to hereafter as the "colour task" and the "shape task"). The required task was indicated to participants with a written cue printed in English upper case and presented centrally on the screen for 500ms. For the colour task, participants needed to determine whether a presented shape was green or blue and indicate their decision by means of a keyboard button press. For the shape task, participants needed to determine whether a presented shape was a square or rectangle. Correct responses were made by means of an 'f' key press for a blue target stimulus in colour trials and a square in shape trials. Correct responses were made by means of a 'j' key press for a green target stimulus in colour trials and a rectangle in shape trials. Participants received feedback on their responses, with a green 'thumbs up' symbol shown for correct responses, and a grey 'thumbs down' symbol for incorrect responses. Feedback was shown for 200ms in the lower portion of the screen, directly below the target stimulus. Participants completed a total of 64 trials, with the presentation of each trial randomised for each participant. Outcome measures for this task included reaction time (total time taken to complete the task), number of total errors, number of errors on colour trials, and number of errors on shape trials.

An online version of the Wisconsin Card Sorting Task (WCST; see Berg, 1948; Grant & Berg, 1948) was used as the final measure of cognitive flexibility. Participants were tasked with matching a target card to one of four stimulus cards presented in the upper portion of the screen. Using standard WCST procedure, participants were not informed of the rules for matching the cards (number, colour, or shape) or of how often these rules would change. Instead, participants are forced to determine the new rule based upon feedback received on the previous attempt. Feedback was again provided by means of a green 'thumbs up' symbol shown for correct responses, and a grey 'thumbs down' symbol for incorrect responses, presented in the lower righthand corner of the screen. Feedback was presented for 700ms. Participants completed 64 trials, with trials randomised within blocks. Outcome measures for this task included reaction time (total time taken to complete the task), number of total errors, number of non-perseverative errors, and number of perseverative errors.

Procedure

The study was administered online using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020; www.gorilla.sc). Data were collected between 24 Nov 2020 and 17 Feb 2021.

Participants were first provided with an information sheet detailing the aims of the study. Those providing their consent to participate were then asked to provide their age, gender (male, female, other), and level of education (doctoral degree, postgraduate degree, undergraduate degree, post-secondary, secondary, vocational). Respondents were also offered an option to not disclose these demographic details. The demographic section of the study asked participants to self-identify as a paranormal believer, a sceptic, or unsure. They were then presented with the CFI, SPQ-B, and the PSBS-14, followed by the CTST and the

WCST. Task order was counter-balanced across participants, so that 50% of the sample completed the CTST first and the remaining 50% completed the WCST first. Before ending the study, participants were presented with a debrief screen.

Informed consent was obtained from all participants and all methods were performed in accordance with relevant guidelines and regulations. The study received ethical approval from the University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (HSET ECDA; protocol number LMS/PGR/UH/04313). The study was preregistered, and the protocol can be found on the Open Science Framework (OSF; https://osf.io/h3vws/).

RESULTS

Relationships between paranormal belief, schizotypy, and the three measures of cognitive flexibility were analysed using R version 4.0.2 (R Core Team, 2020). Significance was set at a level of .05 when calculating *p*-values. Self-assessment measures were coded so that higher scores represented higher levels of the attribute. Descriptive statistics for the self-assessment measures, and the behavioural tests, are reported in Table 4.9.

Table 4.9. Descriptive statistics for all variables in Study 4.2. (*N* = 123).

Measure	*M*	*SD*	Min.	Max.	α
PSBS-14	54.77	20.00	14.00	98.00	.93
SPQ-B total	65.03	14.33	32.00	100.00	.88
SPQ-B interpersonal	23.15	5.48	10.00	37.00	.70
SPQ-B cognitive perceptual	23.96	5.58	11.00	36.00	.71
SPQ-B disorganisation	17.93	4.60	8.00	28.00	.67
CFI total	100.14	14.24	46.00	134.00	.86
CFI alternatives	70.43	11.47	16.00	91.00	.92
CFI control	29.71	6.46	12.00	43.00	.66
WCST total errors	24.65	8.87	9.00	51.00	-
WCST perseverative errors	10.68	7.76	0.00	36.00	-
WCST non-perseverative errors	14.08	3.17	9.00	28.00	-
WCST reaction time	145751.50	48519.22	72092.35	365862.00	-
CTST total errors	10.82	10.95	0.00	38.00	-
CTST colour errors	4.58	5.86	0.00	30.00	-

CTST shape errors	6.24	6.08	0.00	23.00	-
CTST reaction time	66816.81	35968.12	15076.00	236577.80	-

Note: Reaction times measured in milliseconds. PSBS-14 = Paranormal and Supernatural Beliefs Scale 14-item, SPQ-B = Schizotypal Personality Questionnaire Brief, CFI = Cognitive Flexibility Inventory, WCST = Wisconsin Card Sorting Test, CTST = Cued Task-Switching Task.

Correlations between variables

As can be seen in Table 4.10., weak positive correlations emerged between paranormal beliefs and all measures of schizotypal thinking. Weak negative correlations were also found between paranormal belief scores and both WCST reaction times and non-perseverative WCTS errors. All remaining relationships with paranormal beliefs were non-significant at the .05 level.

Table 4.10. Pearson's product-moment correlations to paranormal beliefs.

Measure	*r*	*p*
SPQ-B total	**.36**	**< .001**
SPQ-B interpersonal	**.33**	**< .001**
SPQ-B cognitive perceptual	**.26**	**.004**
SPQ-B disorganisation	**.41**	**< .001**
CFI total	-.04	.67
CFI alternatives	.01	.95
CFI control	-.10	.29
WCST total errors	-.10	.25
WCST perseverative errors	-.03	.73
WCST non-perseverative errors	**-.22**	**.01**
WCST reaction time	**-.23**	**.01**
CTST total errors	.07	.48
CTST colour errors	.11	.24
CTST shape errors	.01	.88
CTST reaction time	-.08	.36

Note: Significant correlations ($p < .05$) presented in bold. SPQ-B = Schizotypal Personality Questionnaire Brief, CFI = Cognitive Flexibility Inventory, WCST = Wisconsin Card Sorting Test, CTST = Cued Task-Switching Task.

When exploring relationships to schizotypal thinking, Table 4.11. shows weak negative correlations between total SPQ-B score and total CFI score, and between SPQ-B interpersonal scores and both total WCST errors and perseverative WCST errors. A moderate negative correlation was found between schizotypy and CFI Control scores.

Table 4.11. Pearson's product-moment correlations to schizotypy measures.

Measure	**SPQ-B Total**	**SPQ-B Interp.**	**SPQ-B Cog. Perc.**	**SPQ-B Disorg.**
PSBS-14	**.36 (<.001)**	**.33 (<.001)**	**.26 (.004)**	**.41 (<.001)**
SPQ-B total	-	**.91 (<.001)**	**.93 (<.001)**	**.91 (<.001)**
SPQ-B interpersonal	**.91 (<.001)**	-	**.75 (<.001)**	**.73 (<.001)**
SPQ-B cognitive perceptual	**.93 (<.001)**	**.75 (<.001)**	-	**.79 (<.001)**
SPQ-B disorganisation	**.91 (<.001)**	**.73 (<.001)**	**.79 (<.001)**	-
CFI total	**-.31 (<.001)**	**-.35 (<.001)**	**-.26 (.003)**	**-.24 (.008)**
CFI alternatives	-.16 (.08)	-.17 (.07)	-.14 (.13)	-.12 (.18)
CFI control	**-.41 (<.001)**	**-.48 (<.001)**	**-.34 (<.001)**	**-.31 (<.001)**
WCST total errors	-.14 (.12)	**-.20 (.025)**	-.06 (.54)	-.13 (.15)
WCST perseverative errors	-.17 (.07)	**-.22 (.01)**	-.08 (.39)	-.16 (.09)
WCST non-perseverative errors	.01 (.87)	-.01 (.88)	.03 (.72)	.02 (.80)
WCST reaction time	-.02 (.84)	-.07 (.44)	.05 (.55)	-.04 (.65)
CTST total errors	-.04 (.63)	-.06 (.54)	.009 (.92)	-.08 (.37)
CTST colour errors	-.004 (.97)	-.02 (.86)	.04 (.64)	-.04 (.62)
CTST shape errors	-.08 (.40)	-.09 (.34)	-.03 (.78)	-.10 (.25)
CTST reaction time	-.04 (.63)	-.06 (.51)	-.16 (.87)	-.05 (.61)

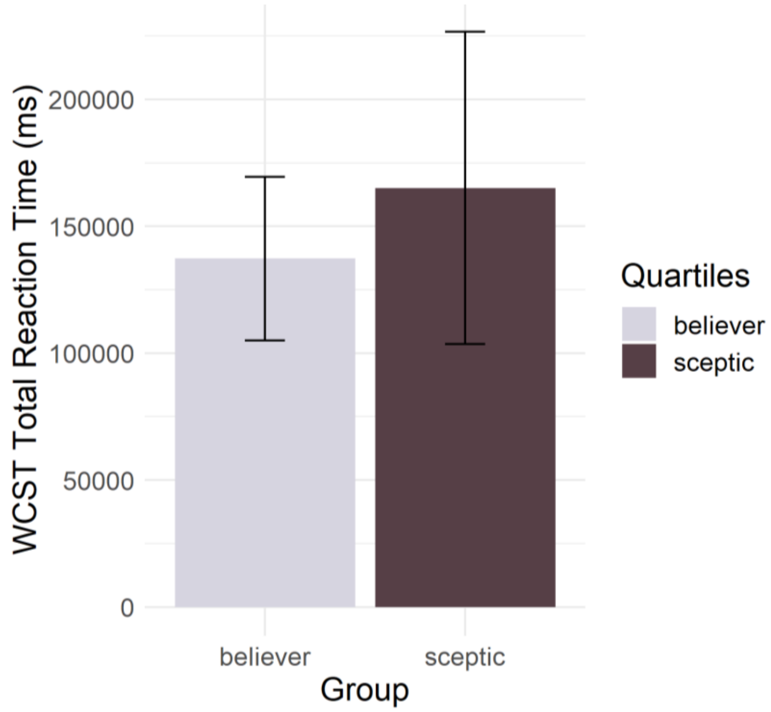
Note: Significant correlations presented in bold, p-values presented in parentheses. PSBS-14 = Paranormal and Supernatural Beliefs Scale 14-item, SPQ-B = Schizotypal Personality Questionnaire Brief, CFI = Cognitive Flexibility Inventory, WCST = Wisconsin Card Sorting Test, CTST = Cued Task-Switching Task.

Between-groups comparisons

Owing to the significant correlations between paranormal beliefs and both reaction times and number of non-perseverative errors on the WCST, between-groups comparisons were made between those scoring high and low on the PSBS-14. Participants scoring above the mean of 54.77 were classified as 'believers' (*n* = 69), while those scoring below were classified as 'sceptics' (n = 51). No significant differences emerged between believers and sceptics for either WCST reaction times or number of non-perseverative errors ($ps > .10$).

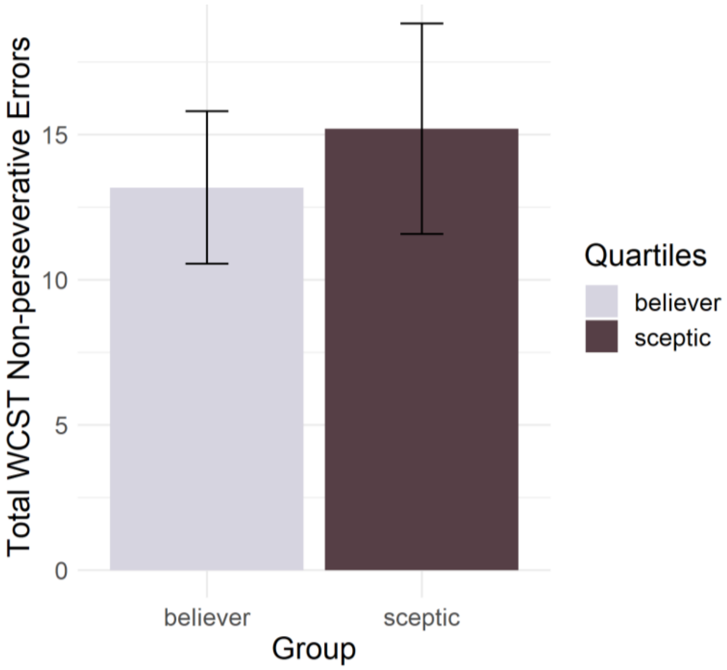
Exploratory analyses then considered those in the lower and upper quartiles, representing the most sceptical participants and those with the strongest beliefs (respectively). Therefore, the following analyses were conducted on a reduced sample of 61 participants (30 believers and 31 sceptics). A significant difference emerged between those in the upper and lower quartiles for reaction times ($t(45.66) = -2.22$, $p = .03$), with believers responding significantly faster than sceptics (M = 137275.60ms, SD = 32255.69ms vs. M = 165054.10ms, SD = 61507.09ms). There was also a significant difference in the number of non-perseverative errors made in the WCST by those in the upper and lower quartiles ($t(59) = -2.50$, $p = .01$), with believers making significantly fewer errors compared to sceptics (M = 13.17, SD = 2.63 vs. M = 15.19, SD = 3.62). The reaction time and non-perseverative error differences can be seen in Figures 4.8. and 4.9., respectively.

Figure 4.8. Mean WCST reaction times in Study 4.2. for the most sceptical participants and those with the strongest beliefs.



Note: WCST = Wisconsin Card Sorting Test. Error bars represent the standard deviations for each group.

Figure 4.9. Mean number of non-perseverative WCST errors in Study 4.2. for the most sceptical participants and those with the strongest beliefs.



Note: WCST = Wisconsin Card Sorting Test. Error bars represent the standard deviations for each group.

Regression and mediation

A multiple linear regression was conducted for paranormal beliefs with the following predictor variables: total SPQ-B score, total CFI score, total CTST errors, total CTST reaction time, total WCST reaction time, total WCST errors, WCST perseverative errors, and WCST non-perseverative errors. A backward stepwise method was then employed to find the most appropriate predictive model for paranormal beliefs, and the results are presented in Table 4.12. It should be noted here that while the preregistration for this study specified the regression would use a stopping rule of $p < .05$, analyses subsequently used the Akaike Information Criterion (AIC) as a stopping rule. This decision was made owing to the inflated risk of Type I errors associated with *p*-value-based model building, and the AIC's strength in preventing over-fitting and providing an estimate of the predictive accuracy of a given model (Halsey, 2019). The algorithm then tries to remove one of the predictor variables from the model to see if there is a change in the AIC value. The variable that minimises the AIC variable the most is removed from the model, with the process repeating until no other variables meet the elimination criteria (i.e., removing subsequent variables would *increase* the AIC value of the model, rather than minimise the value). The purpose of the analysis is to produce a model with the smallest AIC value. However, the differences in AIC (represented hereafter as ΔAIC) are used to compare candidate models and evaluate statistical support for the final model. In cases where the AIC values of the final model and the preceding candidate models differ by less than two, they will be considered to have similar statistical support (Burnham & Anderson, 2002).

Table 4.12. Results of the backward stepwise regression using Akaike information criterion (AIC).

Step	AIC (AIC if variable removed)	Comparison	ΔAIC
1:	**704.26**	N/A	N/A
CTST total errors	(702.26)		
WCST perseverative errors	(702.27)		
WCST total errors	(702.31)		
CTST total RT	(702.53)		
WCST non-perseverative errors	(702.81)		
CFI Total	(703.90)		
WCST total RT	(707.90)		
SPQ-B total	(723.56)		
2:	**702.26**	**Step 1 vs Step 2**	**-2.00**
WCST perseverative errors	(700.27)		
WCST total errors	(700.32)		
CTST total RT	(700.53)		
WCST non-perseverative errors	(700.84)		
CFI Total	(701.90)		
WCST total RT	(705.55)		
SPQ-B total	(721.56)		

3:	**700.27**	**Step 2 vs Step 3**	**-1.99**
CTST total RT	(698.53)		
CFI total	(699.95)		
WCST total errors	(700.36)		
WCST total RT	(703.55)		
WCST non-perseverative errors	(704.91)		
SPQ-B total	(719.56)		
4:	**698.53**	**Step 3 vs Step 4**	**-1.74**
CFI total	(698.08)		
WCST total errors	(698.50)		
WCST total RT	(702.05)		
WCST non-perseverative errors	(703.27)		
SPQ-B total	(717.57)		
5:	**698.08**	**Step 4 vs Step 5**	**-.45**
WCTS total errors	(697.81)		
WCST total RT	(701.15)		
WCST non-perseverative errors	(702.55)		
SPQ-B total	(715.64)		
6:	**697.81**	**Step 5 vs Step 6**	**-.27**
WCST total RT	(700.07)		
WCST non-perseverative errors	(700.55)		
SPQ-B total	(713.87)		

Note: ΔAIC = difference in Akaike information criterion, SPQ-B = Schizotypal Personality Questionnaire Brief, CFI = Cognitive Flexibility Inventory, WCST = Wisconsin Card Sorting Test, CTST = Cued Task-Switching Task.

The final model selected by the backward stepwise regression included total number of non-perseverative errors made on the WCST, total reaction time on the WCST, and total SPQ-B score. Regression coefficients for these variables can be found in Table 4.13. The AIC value of the final model was more than 2 AIC units lower than the AIC value of the full model (ΔAIC = -06.45) suggesting stronger statistical support for the final model over the full model. The final model also had stronger statistical support over the models presented in the second (ΔAIC = -04.45) and third step (ΔAIC = -02.46). However, the models presented in steps four (ΔAIC = -00.72) and five (ΔAIC = -00.27) had similar statistical support to that of the final model. However, for the sake of parsimony, we present the final model as the best-fitting model for the data, as it contains the fewest parameters.

Table 4.13. Final regression model for paranormal beliefs (regression coefficients).

Predictor variable	Regression coefficient
WCST non-perseverative errors	-1.16346
SPQ-B total	0.50029
WCST total RT	-0.00007

Note: SPQ-B = Schizotypal Personality Questionnaire Brief, WCST = Wisconsin Card Sorting Test

Finally, as there is some evidence to suggest that paranormal beliefs mediate the relationship between schizotypal thinking and cognition (see Dagnall et al., 2016), mediation analyses were conducted using the Mediation package (Tingley et al., 2014) for the significant relationships found between schizotypy and cognitive flexibility. For the relationship between total SPQ-B scores and total CFI scores, the bootstrapped unstandardised indirect effect was .03, 95% confidence interval (CI) [.04, .10]. The indirect effect was not statistically significant (p = .384). The bootstrapped unstandardised indirect effect for the relationship between total SPQ-B scores and CFI control scores was .009, 95% CI [-.02, .04]. The indirect effect was not statistically significant (p = .570). For the relationship between SPQ-B interpersonal scores and total number of WCST errors, the bootstrapped unstandardised indirect effect was -.02, 95% CI [-.13, .10]. The indirect effect was not statistically significant (p = .710). Finally, the relationship between SPQ-B interpersonal scores and number of perseverative errors on the WCST, the bootstrapped unstandardised indirect effect was .02, 95% CI [-.07, .15]. The indirect effect was not statistically significant (p = .666). Paranormal beliefs did not mediate any of the relationships found between schizotypal thinking and cognitive flexibility.

DISCUSSION OF STUDY 4.2.

In contrast to the predicted (and previous) findings, Study 4.2. identified a relationship between paranormal beliefs and *increased* cognitive flexibility, with believers demonstrating faster response times and fewer non-perseverative errors on the WSCT. It should be noted, however, that paranormal beliefs and perseverative errors (typically considered to be the primary indication of reduced cognitive flexibility) were not related. There was also no significant relationship between paranormal beliefs and either self-rated cognitive flexibility or any of the CTST outcomes. Schizotypal thinking was again linked to both stronger paranormal beliefs and greater self-rated executive difficulties, with the interpersonal subscale of the SPQ-B also related to increased perseverative errors. While previous research suggested a mediating effect of paranormal beliefs on the relationship between schizotypal thinking and cognition, such an effect was not found in the present study.

The positive association between paranormal beliefs and cognitive flexibility identified here presents an unexpected and interesting finding. Several studies have suggested that reduced cognitive flexibility is associated with an intolerance of uncertainty (e.g., Clarke & Kiropoulos, 2021; Demirtas & Yildiz, 2019; Lieberman et al., 2016), such that individuals with a high intolerance of uncertainty have difficulty switching between cognitive representations. Although a positive relationship between paranormal beliefs and intolerance of uncertainty has been noted by a few studies (e.g., Mauzay & Cuttler, 2018; Hart et al., 2013), this would contradict the present findings if a negative association existed between cognitive flexibility and intolerance of uncertainty. For this explanation to be applicable to the present findings, a positive association would need to exist between intolerance of uncertainty and paranormal *dis*belief. Such a relationship has not been considered in the literature. Some consideration, however, has been given to the relationship between intolerance of uncertainty and dogmatic atheism (i.e., strong, firm religious disbelief), with findings suggesting that those who dogmatically do not believe in religion do so as a cognitive response to uncertainty in much the same way as those who dogmatically believe in religion (Kossowska et al., 2017). Put more simply, both strong religious disbelief and strong religious beliefs help to provide a sense of certainty. As paranormal phenomena, by their very definition, evoke a level of uncertainty, it may be the case that both paranormal beliefs and scepticism share similar characteristics that act to reduce uncertainty.

The findings from Study 4.2. were unexpected and opposed the previous findings described in this chapter, provoking the need for a replication study to confirm these findings. Replicating the present study would also afford the opportunity to extend this further and test for a relationship between scepticism and increased intolerance of uncertainty.

STUDY 4.3. PARANORMAL BELIEFS AND COGNITIVE FLEXIBILITY: REPLICATION AND EXTENSION

The aim of Study 4.3. was to address the conflicting findings identified in Study 4.2. which suggested *increased* cognitive flexibility in paranormal believers (compared to the *reduced* flexibility seen in Study 4.1. and Lindeman et al.'s, 2011, study). The study sought to replicate the findings from the adequately powered sample in Study 4.2., using the same WCST task and a self-report measure of cognitive rigidity. It was hypothesised that paranormal believers would demonstrate increased cognitive flexibility compared to sceptics (demonstrated by faster response latencies and fewer errors on the WCST, as well as lower scores on the cognitive rigidity measure). The study also sought to test the theory that reduced cognitive flexibility may be related to an intolerance of uncertainty. Therefore, based on the findings from Study 4.2., it was predicted that sceptics would score higher on a measure of intolerance of uncertainty than believers.

METHOD

Participants

One hundred and twenty participants were recruited for the replication study using Prolific (https://www.prolific.co/), an online participant recruitment platform. As a-priori power analysis had indicated that a minimum of 123 participants would be needed to detect a medium effect with 90% at an alpha of .05 for the first study, the replication aimed to collect the same number of participants as that achieved in the first study (120). Prior to analysis, five participants were removed due to outlying reaction time data (> 3 standard deviations above the mean). The final sample therefore consisted of 115 participants (53 males, 62 females, aged between 18 and 55; $M = 27.87$, $SD = 8.46$). Of the final sample, the majority (35%) had an undergraduate level of education, followed by a postgraduate education (23%), secondary education (17%), post-secondary education (16%), and doctoral education (4%), with six participants (5%) choosing not to disclose this demographic information.

Materials

Questionnaires

Paranormal beliefs were assessed using the Paranormal and Supernatural Beliefs Scale (PSBS; Dean et al., 2021). This scale was developed using modern test theory methods and presents a more refined and valid assessment of paranormal beliefs compared to the classically developed PSBS-14 that was available for use in Study 4.2. (see Dean et al., 2021, for a detailed comparison of the two scales). Participants rated their level of agreement with 13 items (e.g., "it is possible to become possessed by an evil supernatural entity", "some health conditions can be treated with psychic healing") using a 4-point Likert scale. Total scores could range from 0-39.

Participants' intolerance towards uncertainty was measured using the English version of the Intolerance of Uncertainty Scale (IUS; Buhr & Dugas, 2002; Freeston et al., 1994, see Appendix Q). Participants rated their level of agreement with items such as "uncertainty makes me uneasy, anxious, or stressed" using a 5-point Likert scale (ranging from 1= 'not at all characteristic of me', to 5= 'entirely characteristic of me'). Items cover four domains (uncertainty leads to the inability to act, uncertainty is stressful and upsetting, unexpected events are negative and should be avoided, being uncertain about the future is unfair), and scores on the 27-item scale can range from 27-135. The English translation of the IUS has demonstrated excellent internal consistency (α = .91; Buhr & Dugas, 2002).

The 24-item Detail and Flexibility Questionnaire (DFlex; Roberts et al., 2011, see Appendix R) was used to measure participants' self-assessment of their own cognitive rigidity, and again provide an opportunity to explore potential differences in participants' self-reported versus behavioural cognitive flexibility (or lack thereof). Participants rated their level of agreement with items (e.g., "When others suggest a new way of doing things, I get upset or unsettled") using a 6-point Likert scale. As well as providing a total score, the DFlex provides scores on two subscales: cognitive rigidity (12 items), and attention to detail (12 items). The DFlex demonstrates excellent internal consistency (α = .95; Roberts et al., 2011).

Behavioural tests

The same online WCST used in Study 4.2. was used for the current study with no amendments made to the stimuli or task procedure.

Procedure

The study was once again administered online using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020; www.gorilla.sc). All data were collected on the 22 Jul 2021. Participants were first provided with an information sheet detailing the aims of the study.

Those providing their consent to participate in the study were then asked to provide the same demographic details as those from Study 4.2. Participants were then presented with the DFlex, IUS, and the PSBS. Following the questionnaires, participants were presented with the WCST. Finally, participants were presented with a debrief screen and were automatically redirected back to the Prolific platform following completion of the study for compensation (£3.75).

Informed consent was obtained from all participants and all methods were performed in accordance with relevant guidelines and regulations. The study received ethical approval from the University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (HSET ECDA; protocol number LMS/PGT/UH/04652). The study was preregistered, and the protocol can be found alongside the registration for Study 4.2. on the Open Science Framework (OSF; https://osf.io/h3vws/).

RESULTS

Relationships between paranormal belief, intolerance of uncertainty, cognitive rigidity, and cognitive flexibility were analysed using R version 4.0.2 (R Core Team, 2020). The analyses used the same significance level to that used in study 1 ($p < .05$). Self-assessment measures were again coded so that higher scores represented higher levels of the attribute. Descriptive statistics for each measure are reported in Table 4.14.

Table 4.14. Descriptive statistics for all variables in Study 4.3.

Measure	*M*	*SD*	Min.	Max.	α
PSBS	17.17	7.79	1.00	32.00	.89
IUS total	68.66	20.41	28.00	125.00	.94
DFlex total	83.57	14.21	51.00	117.00	.82
DFlex cognitive rigidity	43.83	7.87	24.00	63.00	.72
DFlex attention to detail	39.74	8.08	20.00	64.00	.73
WCST Total Errors	23.87	9.84	7.00	48.00	-
WCST Perseverative Errors	10.97	8.50	0.00	33.00	-

WCST Non-perseverative Errors	12.87	3.39	7.00	34.00	-
WCST Reaction Time	155061.40	52107.62	74132.40	309390.00	-

Note: Reaction times measured in milliseconds. PSBS = Paranormal and Supernatural Beliefs Scale, IUS = Intolerance of Uncertainty Scale, DFlex = Detail and Flexibility Questionnaire, WCST = Wisconsin Card Sorting Test.

Correlations between variables

As shown in Table 4.15., there were weak positive correlations between paranormal beliefs and the DFlex attention to detail subscale, as well as weak positive correlations between paranormal beliefs and: reaction times, perseverative errors, and number of total errors on the WCST. All remaining relationships with paranormal beliefs were non-significant at the .05 level.

Table 4.15. Pearson's product-moment correlations to paranormal beliefs.

Measure	*r*	*p*
DFlex total	.11	.232
DFlex cognitive rigidity	.00	.963
DFlex attention to detail	**.19**	**.038**
IUS	.05	.590
WCST total errors	**.22**	**.017**
WCST perseverative errors	**.20**	**.030**
WCST non-perseverative errors	.14	.141
WCST reaction time	**.37**	**<.001**

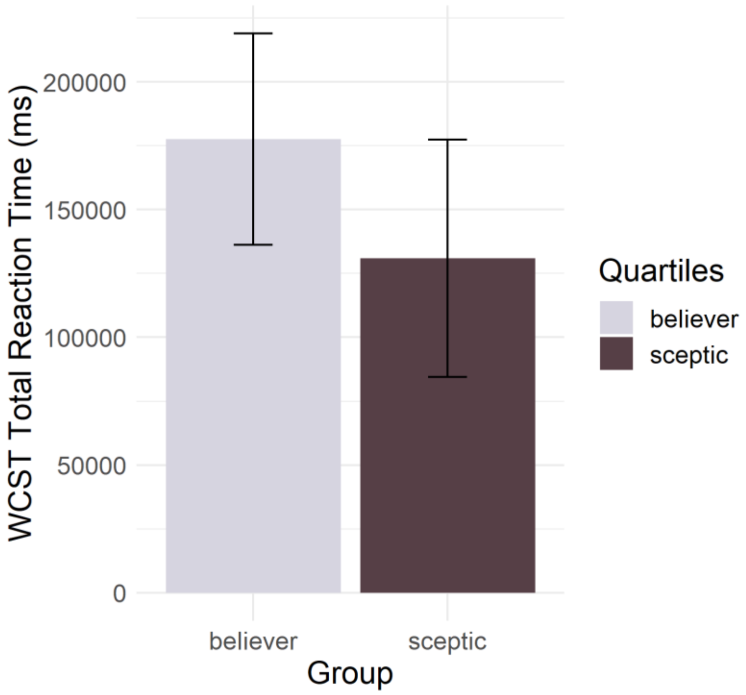
Note: Significant correlations (p < .05) presented in bold. DFlex = Detail and Flexibility Questionnaire, IUS = Intolerance of Uncertainty Scale, WCST = Wisconsin Card Sorting Test.

Between-groups comparisons

To further assess the significant relationships between paranormal beliefs and performance on the WCST, between-groups comparisons were made between those scoring high and low on the PSBS. Participants scoring above the mean of 17.17 were classified as 'believers' (*n* = 65), while those scoring below were classified as 'sceptics' (*n* = 50). Sceptics responded significantly faster than believers (*M* = 135192.70ms, *SD* = 45289.78ms, vs. *M* = 170344.90ms, *SD* = 52157.35ms: $t(113) = -3.79$, $p < .001$). A significant difference also emerged between believers and sceptics for total number of errors on the WCST ($t(113) = -2.57$, $p = .01$), with sceptics making significantly fewer errors than believers (*M* = 21.24, *SD* = 09.51, vs. *M* = 25.89, *SD* = 09.68). Finally, a significant difference emerged between believers and sceptics for perseverative errors on the WCST ($t(113) = -2.24$, $p = .03$), with sceptics making significantly fewer perseverative errors than believers (*M* = 8.98, *SD* = 8.26, vs. *M* = 12.51, *SD* = 8.43).

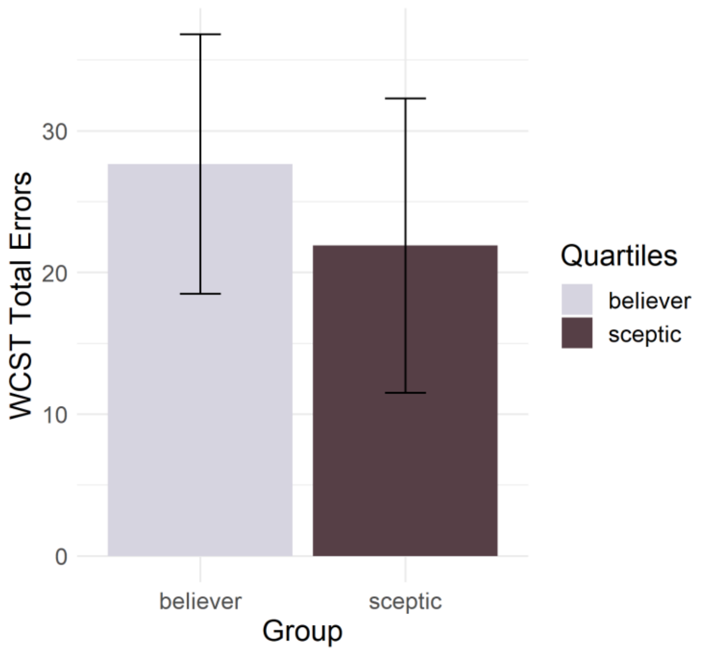
To make the group differences comparable to those from Study 4.2., analyses again focused on the lower and upper quartiles representing the most sceptical participants and those with the strongest beliefs (respectively). The following analyses were therefore conducted on a reduced sample of 67 participants (38 believers and 29 sceptics). A significant difference was found for those in the upper and lower quartiles for reaction times ($t(65) = 4.34$, $p < .001$), with sceptics responding significantly faster than believers (*M* = 130914.20ms, *SD* = 46472.08ms vs. *M* = 177607.30ms, *SD* = 41320.77ms). A significant difference also emerged for the total number of errors made on the WCST ($t(65) = 2.41$, $p = .02$), with sceptics making significantly fewer errors compared to believers (*M* = 21.90, *SD* =10.38 vs. *M* = 27.66, *SD* = 9.14). Finally, sceptics made significantly fewer perseverative errors on the WCST compared to believers (*M* = 9.14, *SD* = 8.66 vs. *M* = 13.71, *SD* = 8.08: $t(65) = 2.22$, $p = .03$). The reaction time and error differences for the upper and lower quartiles can be seen in Figures 4.10., 4.11., and 4.12., respectively.

Figure 4.10. Mean WCST reaction times in Study 4.3. for the most sceptical participants and those with the strongest beliefs.



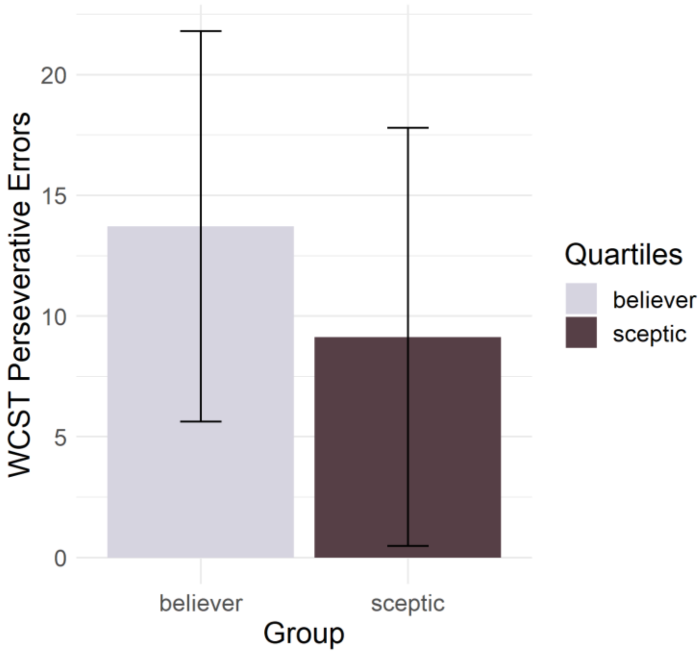
Note: WCST = Wisconsin Card Sorting Test. Error bars represent the standard deviations for each group.

Figure 4.11. Mean number of total WCST errors in Study 4.3. for the most sceptical participants and those with the strongest beliefs.



Note: WCST = Wisconsin Card Sorting Test. Error bars represent the standard deviations for each group.

Figure 4.12. Mean number of WCST perseverative errors in Study 4.3. for the most sceptical participants and those with the strongest beliefs.



Note: WCST = Wisconsin Card Sorting Test. Error bars represent the standard deviations for each group.

Study 4.2. versus Study 4.3.

As Study 4.2. and Study 4.3. produced opposing results on the WCST, exploratory analyses were conducted to assess potential differences in reaction times and errors made in each study. The following analyses were therefore conducted with a combined sample size of 235 participants (120 from Study 4.2. and 115 from Study 4.3.).

For the total samples, no significant differences emerged between Study 4.2. and Study 4.3. for reaction times on the WCST (M = 145751.50ms vs 155061.40ms, $t(233) = -1.42$, $p = .16$), or for the number of total errors made (M = 24.65 vs 23.87, $t(233) = 0.64$, $p = .52$), or the number of perseverative errors (M = 10.68 vs 10.97, $t(233) = -0.28$, $p = .78$). A significant difference was found, however, for the number of non-perseverative errors made in Study 4.2. and Study 4.3. ($t(233) = 2.84$, $p = .005$), with participants in Study 4.2. making significantly more non-perseverative errors (M = 14.08, SD = 3.17) than those in Study 4.3. (M = 12.87, SD = 3.39).

For paranormal believers, there was a significant difference in reaction times in Study 4.2. and Study 4.3. ($t(132) = -3.49, p < .001$). Believers in Study 4.2. were significantly faster on the WCST (M = 141322.90ms, SD = 43912.62ms) than believers in Study 4.3. (M = 170344.90ms, SD = 52157.35ms). There was no significant difference in the number of total errors made by believers in Study 4.2. and Study 4.3. (M = 24.03 vs 25.89, $t(132) = -1.16, p = .25$). There was also no significant difference in the number of perseverative errors made by believers in Study 4.2. and Study 4.3. (M = 10.45 vs 12.51, $t(132) = -1.44, p = .15$), or the number of non-perseverative errors (M = 13.70 vs 13.37, $t(120.70) = 0.54, p = .59$).

Finally, exploratory analyses of sceptics' responses in Study 4.2. and Study 4.3. found no significant difference in reaction times (M = 151743.20ms vs 135192.70ms, $t(99) = 1.67, p = .10$), or in the number of perseverative errors made (M = 10.98 vs 8.98, $t(99) = 1.30, p = .20$). However, there was a significant difference in the total number of errors made by sceptics in Study 4.2. and Study 4.3. ($t(99) = 2.34, p = .02$). Sceptics in Study 4.2. made significantly more errors (M = 25.49, SD = 8.78) than sceptics in Study 4.3. (M = 21.24, SD = 9.51). There was also a significant difference in the number of non-perseverative errors made by sceptics in Study 4.2. and Study 4.3. ($t(92.31) = 4.13, p < .001$). Sceptics in Study 4.2. made significantly more non-perseverative errors (M = 14.61, SD = 3.30) than sceptics in Study 4.3. (M = 12.22, SD = 2.45).

Study 4.2. versus Study 4.3. – education

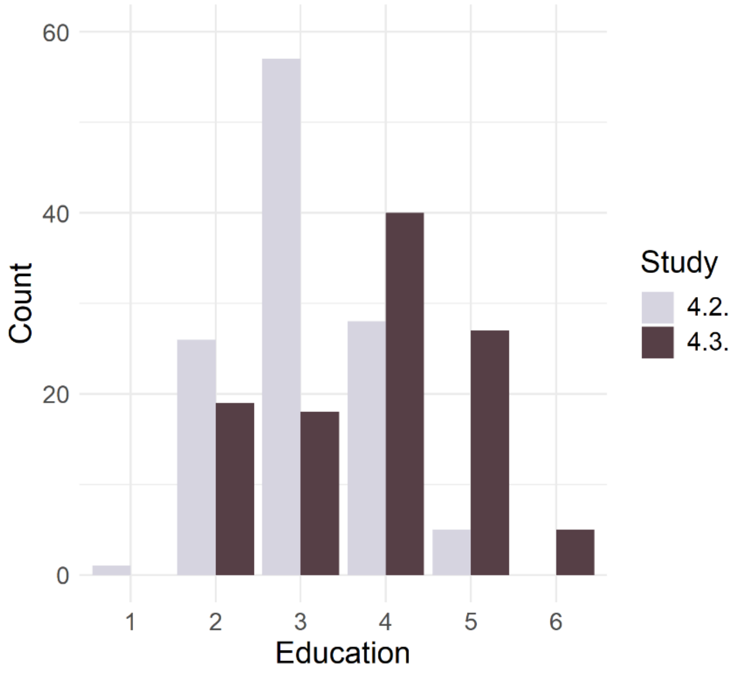
As previous research has suggested a positive relationship between education and cognitive functioning (e.g., Lövdén et al., 2020; Guerra-Carrillo et al., 2017), analyses were conducted to explore participants' level of education in Study 4.2. and Study 4.3. which could account for the differences seen above for WCST performance. Prior to analyses, six participants were removed from the Study 4.3. sample who chose not to disclose information relating to their level of education, and three were removed from the Study 4.2. sample. Therefore, the following analyses were conducted on a slightly reduced sample of 226 participants (117 from Study 4.2. and 109 from Study 4.3.). Education categories were collapsed into two categories: (1) degree (undergraduate degree, postgraduate degree, doctoral degree), and (2) no degree (vocational education, secondary education, post-secondary education). A Pearson's Chi-squared test with Yates' continuity correction revealed no significant association between paranormal beliefs and education in Study 4.2. (χ^2 (1, N = 117) = 1.91, p = .17), and no significant association between paranormal beliefs and education in Study 2.3. (χ^2 (1, N = 109) = 0.75, p = .39). There was a significant association, however, between education and study (χ^2 (1, N = 226) = 31.00, $p < .001$), with an overall higher level of education in Study 4.3. Table 4.16. shows the number of participants in the "degree" and "no degree" categories across both studies, with Figure 4.13. displaying the number of participants in each original education category.

Table 4.16. Education level of participants in Study 4.2. versus Study 4.3.

Study	Degree	No Degree
4.2.	33	84
4.3.	72	37

Note: Degree = undergraduate degree, postgraduate degree, doctoral degree; No Degree = vocational education, secondary education, post-secondary education

Figure 4.13. Participants' level of education in Study 4.2. versus Study 4.3.



Note: Education categories – 1 = vocational education, 2 = secondary education, 3 = post-secondary education, 4 = undergraduate education, 5 = postgraduate education, 6 = doctoral education.

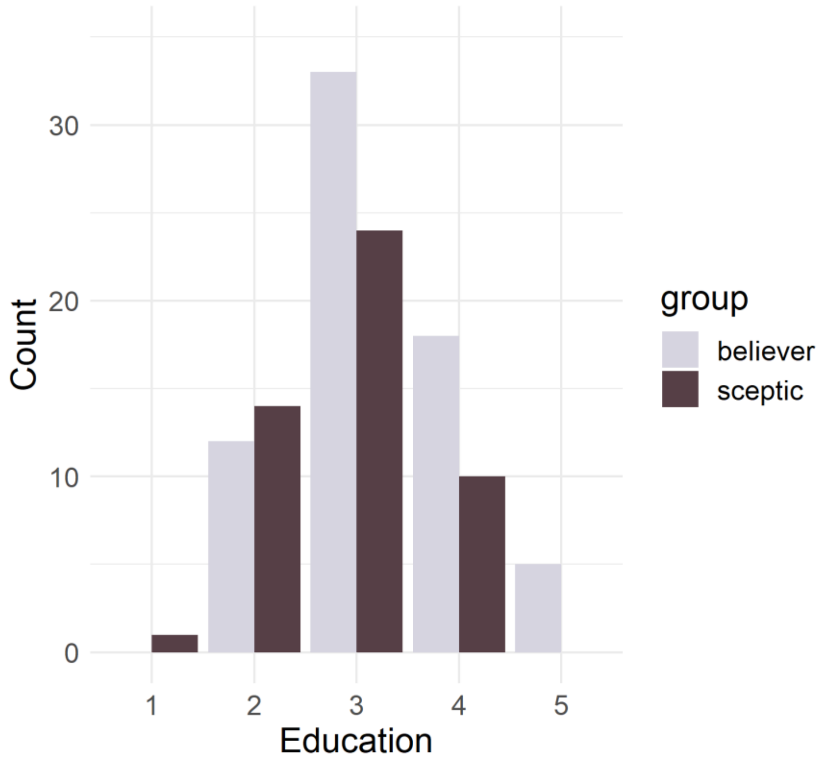
Believers had a higher level of education compared to sceptics in Study 4.2., although this difference was not statistically significant. Table 4.17. shows the number of believers and sceptics in the "degree" and "no degree" categories across both studies. Figure 4.14. displays the number of believers and sceptics in each original education category in Study 4.2., with Figure 4.15. displaying the same information for Study 4.3.

Table 4.17. Education level of believers and sceptics in Study 4.2. versus Study 4.3.

Study	Group	Degree	No Degree
4.2.	Believer	23	46
	Sceptic	10	41
4.3.	Believer	37	23
	Sceptic	35	14

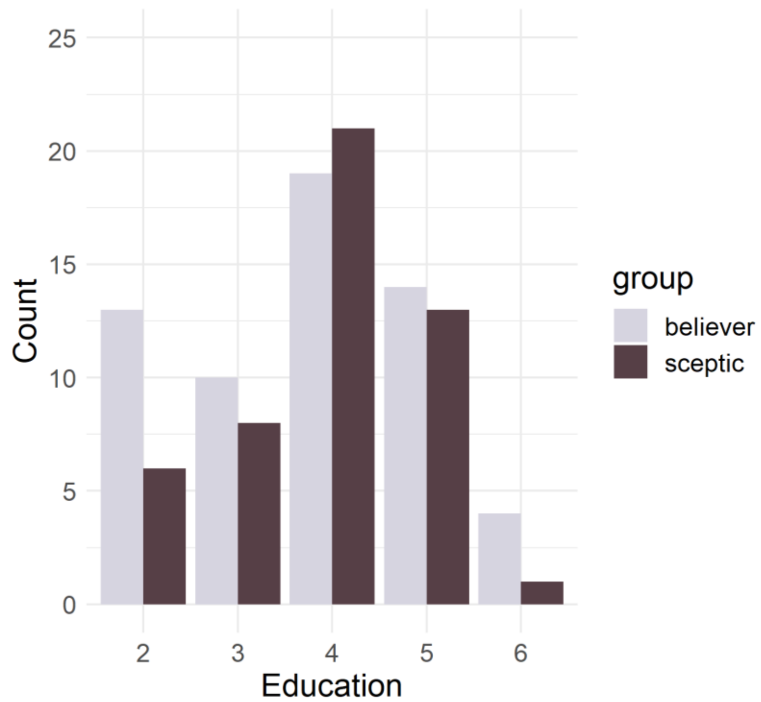
Note: Degree = undergraduate degree, postgraduate degree, doctoral degree; No Degree = vocational education, secondary education, post-secondary education.

Figure 4.14. Believers' and sceptics' level of education in Study 4.2.



Note: Education categories – 1 = vocational education, 2 = secondary education, 3 = post-secondary education, 4 = undergraduate education, 5 = postgraduate education, 6 = doctoral education.

Figure 4.15. Believers' and sceptics' level of education in Study 4.3.



Note: Education categories – 1 = vocational education, 2 = secondary education, 3 = post-secondary education, 4 = undergraduate education, 5 = postgraduate education, 6 = doctoral education.

DISCUSSION OF STUDY 4.3.

Study 4.3. failed to replicate the positive association between cognitive flexibility and paranormal beliefs found in Study 4.2. The findings here suggest a negative relationship, with believers demonstrating slower response times and an increased number of total and perseverative errors on the WSCT compared to sceptics. There was also no significant relationship between paranormal beliefs and intolerance of uncertainty; however, there was a small positive correlation between paranormal beliefs and the DFlex Attention to Detail subscale. Higher scores on this subscale are thought to be indicative of a "weak central coherence" (i.e., a detail-focused processing bias). In this sense, paranormal believers demonstrated a preference for focusing on "parts" versus "wholes". Weaker central coherence has been related to poorer cognitive flexibility (e.g., Danner et al., 2012) and could provide a potential explanation for the increased number of perseverative errors made by believers in the present study.

While the present findings failed to replicate those from Study 4.2., they do provide some support for the findings previously discussed in this chapter. Specifically, the present findings support the negative relationship between paranormal beliefs and cognitive flexibility reported in Study 4.1. and Lindeman et al.'s (2011) work. A possible explanation for the failure to replicate the findings from Study 4.2., despite the present study using the same measure of cognitive flexibility (WCST) and an almost identical sample size to that of Study 4.2., is the significantly higher level of education seen in the sample for Study 4.3. Previous research has shown many advantageous cognitive outcomes of higher education, including increased cognitive functioning and a reduced risk of age-related cognitive impairments

(Nyberg et al., 2021; Lövdén et al., 2020). With a specific focus on executive functioning, studies have demonstrated a positive relationship between the number of years spent in formal education and executive functions such as: cognitive flexibility, attention, planning skills, verbal fluency, and working memory (see Loftus et al., 2021; de Oliveira Souza et al., 2013; Voos et al., 2011; Clark et al., 2004). The failure to replicate the findings from Study 4.2. may therefore be attributable to the significant difference in education between the samples used in Study 4.2. and Study 4.3. This raises the need for an additional replication to determine whether education influences the relationship between paranormal beliefs and cognitive flexibility.

STUDY 4.4. PARANORMAL BELIEFS AND COGNITIVE FLEXIBILITY: REPLICATION AND EXTENSION 2

Owing to the exploratory findings relating to the education level of the previous samples, the aim of Study 4.4. was to explore education as a potential explanation for the conflicting cognitive flexibility findings in Study 4.2. and Study 4.3. The study therefore sought to recruit two groups of participants with differing levels of education. Recruitment for the higher education sample required participants to have a university level of education (undergraduate, postgraduate, or doctoral education), while recruitment for the lower education sample required participants to have lower than a university level of education (vocational, secondary, or post-secondary education). Based on the findings from the previous studies, it was predicted that paranormal believers would demonstrate reduced cognitive flexibility compared to sceptics when the overall level of education of the sample is high, and better cognitive flexibility compared to sceptics when the overall level of education of the sample is low.

METHOD

Participants – Higher education

One hundred and thirty participants were recruited using the online participant recruitment system Prolific (https://www.prolific.co/). Inclusion criteria required that participants had an undergraduate, postgraduate, or doctoral level of education. A-priori power analysis indicated that a minimum of 136 participants would be needed to detect a medium effect with 90% power at an alpha of .05. Prior to analysis, three participants were removed due to outlying reaction time data (> 7 standard deviations above the mean). The final higher education sample therefore consisted of 127 participants (66 males, 59 females, 2 unreported, aged between 18 and 56; $M = 25.91$, $SD = 5.77$).

Participants – Lower education

One hundred and thirty participants were again recruited using the online participant recruitment system Prolific (https://www.prolific.co/). Inclusion criteria required that participants did not have a university level of education (vocational education, secondary education, or post-secondary education). A-priori power analysis indicated that a minimum of 136 participants would be needed to detect a medium effect with 90% power at an alpha of .05. Prior to analysis, one participant was removed due to outlying reaction time data (> 7 standard deviations above the mean). The final higher education sample therefore consisted of 129 participants (79 males, 45 females, 5 unreported, aged between 18 and 56; $M = 22.66$, $SD = 5.61$).

Materials

Questionnaires

Paranormal beliefs were again assessed using the Paranormal and Supernatural Beliefs Scale (PSBS; Dean et al., 2021).

Behavioural tests

The same online WCST used in the previous studies was used for the current study with no amendments made to the stimuli or task procedure.

Procedure

The study was again administered online using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020; www.gorilla.sc). All data were collected on the 13 Feb 2022. Participants were first provided with an information sheet detailing the aims of the study. Those providing their consent to participate in the study were then asked to provide the same demographic details as those from Study 4.2. and Study 4.3. and were presented with the PSBS followed by the WCST. Finally, participants were presented with a debrief screen and were automatically redirected back to the Prolific platform following completion of the study for compensation (£2.50).

Informed consent was obtained from all participants and all methods were performed in accordance with relevant guidelines and regulations. The study received ethical approval from the University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (HSET ECDA: protocol number LMS/PGT/UH/04652(1)). The study was preregistered, and the protocol can be found alongside the registrations for Studies 4.2. and 4.3. on the Open Science Framework (OSF; https://osf.io/h3vws/).

RESULTS

Relationships between paranormal belief and cognitive flexibility were again analysed using R version 4.0.2 (R Core Team, 2020) using the significance level of $p < .05$. Descriptive statistics are reported in Table 4.18.

Table 4.18. Descriptive statistics for all variables in Study 4.4.

Group	Measure	*M*	*SD*	Min.	Max.
Higher education	PSBS	16.54	5.09	2.00	30.00
	WCST Total Errors	20.90	8.58	8.00	53.00
	WCST Perseverative Errors	8.42	7.40	0.00	31.00
	WCST Non-perseverative Errors	12.48	2.58	7.00	25.00
	WCST Reaction Time	131594.54	35908.87	68247.20	243503.00
Lower education	PSBS	14.60	7.70	0.00	34.00
	WCST Total Errors	22.57	9.67	9.00	50.00
	WCST Perseverative Errors	9.60	8.15	0.00	34.00
	WCST Non-perseverative Errors	12.97	3.80	8.00	31.00
	WCST Reaction Time	134024.45	49146.86	60240.60	300209.70

Note: Reaction times measured in milliseconds. PSBS = Paranormal and Supernatural Beliefs Scale, WCST = Wisconsin Card Sorting Test.

Correlations between variables

As shown in Table 4.18., the correlations between paranormal belief and WCST performance for the higher education sample were positive for reaction time and negative for the different error types (total, perseverative, and non-perseverative), suggesting a potential speed-error trade-off. These correlations, however, did not reach statistical significance. Weak positive correlations emerged for the lower education sample between paranormal beliefs and all WCST outcome measures; these relationships were only significant, however, for the number of total errors and perseverative errors made.

Table 4.18. Pearson's product-moment correlations to paranormal beliefs.

Group	Measure	*r*	*p*
Higher education	WCST total errors	-.07	.457
	WCST perseverative errors	-.05	.559
	WCST non-perseverative errors	-.07	.422

	WCST reaction time	.16	.067
Lower education	WCST total errors	**.25**	**.004**
	WCST perseverative errors	**.28**	**.001**
	WCST non-perseverative errors	.04	.644
	WCST reaction time	.11	.225

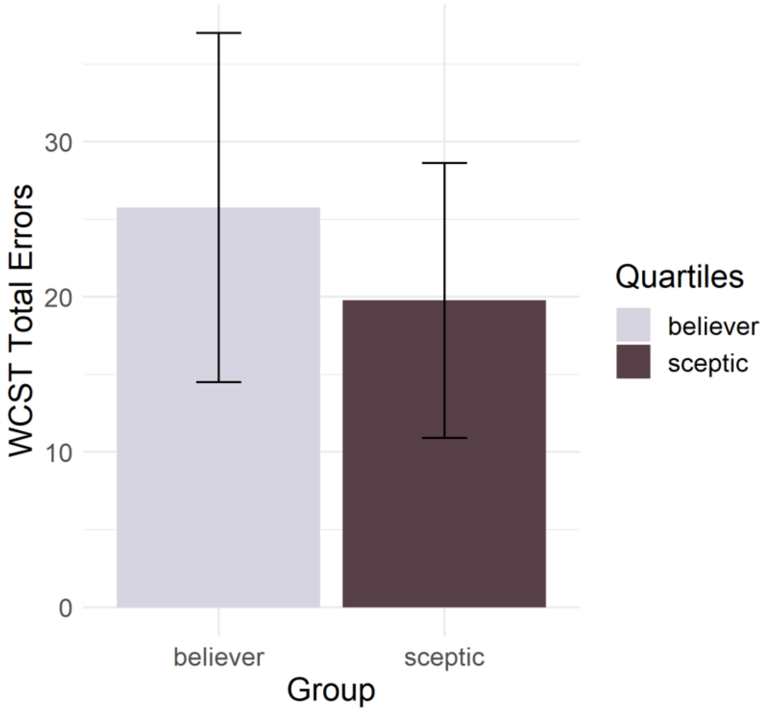
Note: Significant correlations (p < .05) presented in bold. WCST = Wisconsin Card Sorting Test.

Between-groups comparisons

As no relationship was found between paranormal beliefs and WCST performance in the higher education sample, between-groups comparisons were not conducted for this group. Between-groups comparisons were made, however, to further assess the significant relationships found in the lower education sample between paranormal beliefs and both total and perseverative WCST errors. Participants in the lower education sample were divided using the mean total PSBS score for the group, with those scoring above the mean of 14.60 classified as 'believers' (n = 63) and those scoring below classified as 'sceptics' (n = 66). There was a significant difference between believers and sceptics for total errors ($t(127) = -2.90$, $p = .004$), with sceptics making significantly fewer errors compared to believers (M = 20.23, SD = 8.81, vs. M = 25.03, SD = 9.99). There was also a significant difference between believers and sceptics for perseverative errors ($t(116.18) = -3.47$, $p = .001$), with sceptics making significantly fewer perseverative errors than believers (M = 7.26, SD = 6.74, vs. M = 12.06, SD = 8.80).

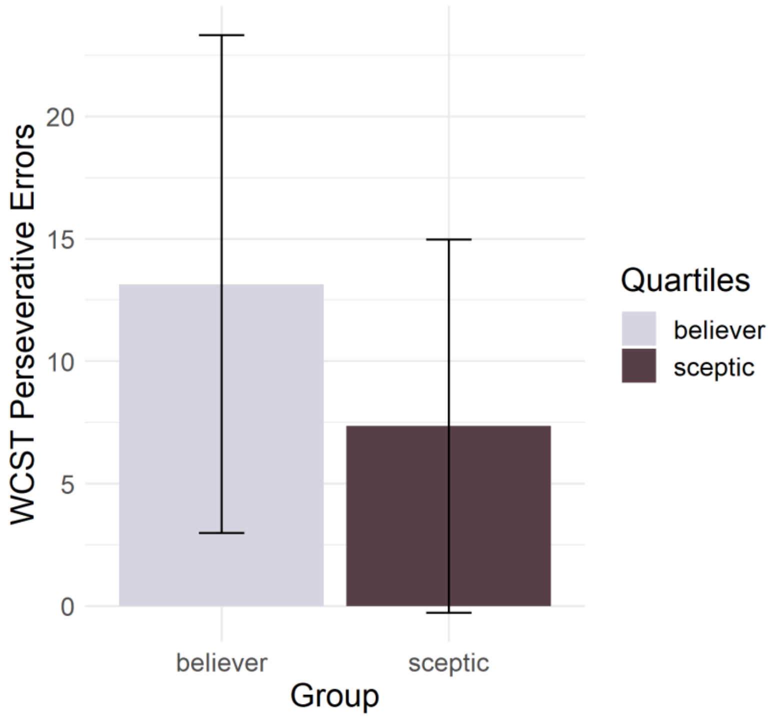
To make the group differences comparable to those from Study 4.2. and Study 4.3., analyses again focused on the lower and upper quartiles representing the most sceptical participants and those with the strongest beliefs in the lower education sample (respectively). The following analyses were therefore conducted on a reduced sample of 69 participants (35 believers and 34 sceptics). A significant difference emerged between believers and sceptics for total errors ($t(67) = 2.45$, $p = .02$), with sceptics making significantly fewer errors compared to believers (*M* = 19.76, *SD* = 8.85, vs. *M* = 25.74, *SD* = 11.24). Sceptics also made significantly fewer perseverative errors compared to believers (*M* = 7.35, *SD* = 7.62, vs. *M* = 13.14, *SD* = 10.17: $t(67) = 2.67$, $p = .009$). The total and perseverative error differences for the upper and lower quartiles in the lower education sample can be seen in Figures 4.16. and 4.17., respectively.

Figure 4.16. Mean number of WCST total errors for the most sceptical participants and those with the strongest beliefs in the lower education sample.



Note: WCST = Wisconsin Card Sorting Test. Error bars represent the standard deviations for each group.

Figure 4.17. Mean number of WCST perseverative errors for the most sceptical participants and those with the strongest beliefs in the lower education sample.



Note: WCST = Wisconsin Card Sorting Test. Error bars represent the standard deviations for each group.

Higher education sample & Lower education sample combined data

Data from both groups of participants (higher education sample and lower education sample) were combined to allow for an exploration of potential main and interaction effects. Specifically, analyses were focused on the main and interaction effects of belief and education on WCST reaction times and total errors, as well as for WCST error type (perseverative versus non-perseverative). The following analyses were conducted with a combined sample size of 256 participants (127 from the higher education sample and 129 from the lower education sample).

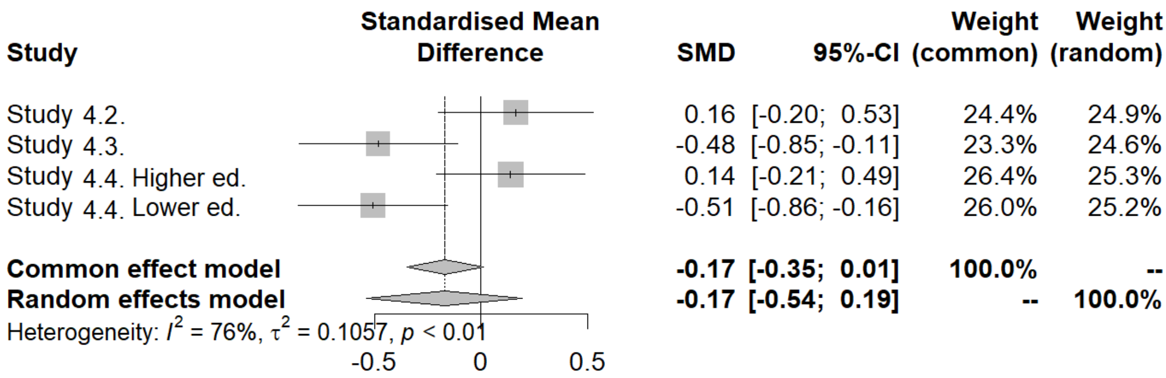
A significant main effect of paranormal beliefs emerged for reaction times for the WCST ($F(1, 252) = 6.73$, $p = .01$), but no significant main effect was identified for education and no interaction effect was identified between paranormal beliefs and education (Fs < 1).

Sceptics had faster reaction times for the WCST compared to believers, regardless of their level of education (*M* = 126420.48ms vs 140301.98ms). For the total number of errors made on the WCST, a significant main effect of paranormal beliefs emerged ($F(1, 252) = 4.58$, $p = .03$). The main effect of education and the interaction between education and paranormal beliefs were not significant (*p*s > .05). Regardless of their level of education, sceptics made fewer total errors on the WCST compared to believers (*M* = 20.60 vs 23.10). Finally, a significant main effect of error type was identified ($F(1, 252) = 57.00$, $p < .001$), with participants making more non-perseverative errors than perseverative errors (*M* = 12.73 vs 09.02) regardless of their level of education or paranormal beliefs. A significant interaction effect emerged between error type and paranormal beliefs, with sceptics making fewer perseverative errors compared to believers (*M* = 7.98 vs 10.23). While sceptics did also make fewer non-perseverative errors compared to believers, this difference was very marginal (*M* = 12.64 vs 12.82). No significant two-way interaction (between error type and level of education emerged), or three-way interaction (between error type, level of education, and paranormal beliefs) effects emerged.

Meta analysis

As a final exploratory analysis conducted in attempt to clarify the relationship between paranormal beliefs and cognitive flexibility, a meta-analysis was conducted for total errors made on the WCST in Studies 4.2., 4.3., and 4.4. There was no overall significant difference in cognitive flexibility between believers and sceptics, with a small, pooled effect size of $g = -.17$, 95% CI [-0.54, 0.19]. While the difference between the groups was not significant, the effect size is suggestive of a negative association between paranormal beliefs and cognitive flexibility. Heterogeneity was high across the studies ($I^2 = 76.00$, $Q = 12.52$, $df = 3$, $p = .006$), with two studies showing more WCST errors made by believers and two showing more WCST errors made by sceptics. The forest plot for this analysis can be found in Figure 4.18. below.

Figure 4.18. Forest plot for total WCST errors made in Studies 4.2., 4.3., and 4.4. as a function of belief group (positive SMD values indicative of positive relationship between paranormal beliefs and cognitive flexibility).



Note: WCST = Wisconsin Card Sorting Test, SMD = Standardised Mean Difference.

DISCUSSION OF STUDY 4.4.

The findings from Study 4.4. provide some additional support for the negative association between paranormal beliefs and cognitive flexibility noted in Study 4.1., Study 4.3., and Lindeman et al.'s (2011) work. When examining participants with higher versus lower levels of education, paranormal believers made significantly more perseverative and total errors on the WCST than sceptics, but only in the lower education sample. Combining the data from both groups highlighted significant effects of paranormal beliefs on WCST performance, with sceptics demonstrating faster response times than believers in addition to fewer total and perseverative errors.

As previously discussed, studies have suggested that higher education is associated with improved executive functioning (see Loftus et al., 2021; de Oliveira Souza et al., 2013; Voos et al., 2011; Clark et al., 2004). While the present findings do not support an effect of education on cognitive flexibility, it is interesting to note that no significant associations were found between paranormal beliefs and cognitive flexibility in the higher education sample. One possibility is that while education itself did not have a direct influence on cognitive flexibility, factors relating to the number of years spent in formal education mediate the relationship between paranormal beliefs and cognitive flexibility. For example, moderate positive associations have been noted between years spent in formal education and fluid intelligence (see Jackson et al., 2020; Gnambs, 2017; Kaufman et al., 2009). Fluid intelligence (described as the ability to adapt to novel situations in which crystallised knowledge offers no advantage; Cattell, 1963) in turn is thought to overlap with several executive functions, particularly cognitive flexibility (Diamond, 2013; Colzato et al., 2006). In line with the fluid-executive theory (Dean et al., 2022), this could explain why a significant relationship was found between paranormal beliefs and cognitive flexibility in the lower education sample but not in the higher education sample, despite no direct main or interaction effects of education. Additional replications of this work would therefore be required to test for a potential mediating effect of fluid intelligence on the relationship between paranormal beliefs and cognitive flexibility.

Replication studies in this area, however, must consider the impact of practice effects on executive tests. As Basso et al. (1999) document, performance on some of the most administered executive tests (including the WCST) changes significantly as a function of repeated administration, such that test scores increase significantly across a 12-month period. Data collected in such studies are, therefore, at risk of being confounded by practice effects if participants have prior experience of the executive test being used. This is particularly important to consider in studies relying on crowdsourcing services (e.g., Prolific, Amazon Mechanical Turk, etc.) for recruitment, as these samples are likely to participate across multiple related experiments (Chandler et al., 2013) and are, therefore, more likely to have completed other executive studies using the same or similar materials. While Studies 4.3. and 4.4. did include some consideration of participants' prior exposure to similar studies (by requesting that they not complete the study if they had participated in Study 4.2.), the possibility of these practice and measurement errors do present a potential limitation of the present work. Future replication studies in this area must, therefore, not only consider potential mediating factors in the relationship between paranormal beliefs and executive function, but also acknowledge the importance of sample naïveté and include pre-screening measures to manage the exclusion of participants with prior executive test experience.

CHAPTER 4 GENERAL DISCUSSION

While the studies presented in this chapter provide some indication that believers and sceptics differ in their cognitive flexibility, the findings are not consistent and point to the influence of additional mediator variables. Studies 4.1. and 4.3. suggested a negative relationship between paranormal beliefs and cognitive flexibility, with sceptics demonstrating faster reaction times and fewer errors than believers. Similar findings were noted in Study 4.4., but only for those with a lower level of education. Study 4.2. presented opposing findings, with believers demonstrating faster response times and fewer errors than sceptics. Meta-analysis of the data from Study 4.2. to 4.4. revealed an overall weak negative effect (g = -.17, 95% CI [-0.54, 0.19]) with high heterogeneity between studies. Attempts to determine potential mediating factors of the relationship between paranormal beliefs and cognitive flexibility considered the roles of intolerance of uncertainty (Study 4.3.) and education (Study 4.4.). Neither factor, however, was associated with paranormal beliefs or cognitive flexibility. Table 4.19. presents a summary of significant WCST performance differences as a function of belief group in Studies 4.2. to 4.4.

Table 4.19. Summary of significant WCST performance differences as a function of belief group in Studies 4.2. to 4.4.

Study	Believers increased performance	Believers reduced performance
4.2.	Reaction times: $t(45.66) = -2.22$, $p = .03$ Non-persev. errors: $t(59) = -2.50$, $p = .01$	–
4.3.	–	Reaction times: $t(65) = 4.34$, $p < .001$ Total errors: $t(65) = 2.41$, $p = .02$ Persev. errors: $t(65) = 2.22$, $p = .03$
4.4. Higher education	–	–
4.4. Lower education	–	Total errors: $t(67) = 2.45$, p .02 Persev. errors: $t(67) = 2.67$, $p = .009$

Note: WCST = Wisconsin Card Sorting Test, Non-persev. = Non-perseverative, Persev. = Perseverative

Paranormal beliefs were found to be weakly related to reduced cognitive flexibility. Nonetheless, the high heterogeneity associated with the pooled estimated means the results from the studies presented in this chapter should be interpreted with some caution. Results of multiple studies will always differ to some degree; however, variation in underlying target parameters becomes a concern when it is substantial and cannot be explained by random variation or identifiable differences in study populations, outcome measures, or methodologies (Imrey, 2020; Alba et al., 2016). While studies 4.2. to 4.4. used the same WCST outcome measures and methodological procedures to assess cognitive flexibility, differences were noted in the studies' populations. Specifically, participants in Study 4.3. had a significantly higher level of education than participants in Study 4.2. Investigation of the potential role of education on the relationship between paranormal beliefs and cognitive flexibility, however, was unable to identify this as a cause of the variation in study findings. Although the same cognitive flexibility measure (WCST) was used across studies 4.2. to 4.4., this task could contribute to the high variability seen across the studies. As previously described, the WCST (like many measures of executive functioning) suffers from the task impurity problem as successful completion of the task requires several cognitive abilities. The variability seen across studies could therefore be attributed to variations in different cognitive abilities between the samples that are recruited for the WCST but were not directly measured. For example, as discussed above, fluid intelligence is thought to overlap with cognitive flexibility. Fluctuations in the levels of fluid intelligence in each study population could therefore confound the relationships identified between paranormal beliefs and cognitive flexibility. This could account for the different findings identified for the higher education and lower education samples in Study 4.4., despite no main or interaction effects of education, as fluid intelligence has shown positive associations with years of formal education. Similarly, the WCST is thought to recruit lexical, numerical, and visual processing abilities (which have also shown positive associations with level of education, see Poreh et al., 2015; De Luccia & Ortiz, 2009; Van Strien et al., 2009), and so variations in any of these cognitive abilities between the study populations could impact performance on the WCST. Future replications of this work would therefore need to consider and measure these additional cognitive abilities to determine their impact on the association between paranormal beliefs and cognitive flexibility.

The nature of the relationship between paranormal beliefs and cognitive flexibility is far from clear. It is clear, however, that this association cannot be described as a "deficit" of executive functioning owing to the heterogeneity between study findings and the weak pooled effect size. Further research is recommended to consider the potential mediating effects of additional cognitive abilities (e.g., fluid intelligence) on the relationship between paranormal beliefs and cognitive flexibility.

CHAPTER 5. A QUALITATIVE APPROACH TO PARANORMAL BELIEFS

INTRODUCTION

Why use qualitative methods for paranormal research?

The aim of qualitative research, when applied in any field, is to illuminate aspects of human life through a lens of subjective meaning and lived experience (Fossey et al., 2002). Where quantitative research strives to make predictions, generalise findings, and determine causality, qualitative research strives for illumination, deeper understanding, and extrapolation of findings to similar situations (Hoepfl, 1997). As Kruth (2015) notes, research in parapsychology frequently makes use of quantitative research methods, but the detailed information derived from qualitative methods is essential to further our knowledge of the field. While the volume of literature is small, several qualitative studies have been conducted to explore the subjective importance of both paranormal beliefs and paranormal experiences.

The function of belief and experience

The adaptive value of paranormal beliefs has not been explored extensively using either qualitative or quantitative methods. Recent work by Betsch et al. (2021), however, employed a semi-structured interview technique to explore this topic with fifteen participants. Through qualitative content analysis, Betsch et al. (2021) found paranormal beliefs to be functional for two main motives: striving for mastery (e.g., goal setting, coping with fear, overcoming life events) and striving for a positive evaluation of the self (e.g., self-awareness, caring for the self). Despite the positive functions paranormal beliefs served, participants also reported negative impacts, citing social isolation and exclusion from social groups as the biggest challenges to their beliefs. However, Betsch et al. (2021) noted that personal experiences of the paranormal helped to reinforce paranormal beliefs when these beliefs were threatened by social challenges. This not only highlights the functional aspect of paranormal beliefs, but also highlights the importance of paranormal experiences for the maintenance of paranormal beliefs. Eaton's (2015) interviews with paranormal investigators (or "ghost hunters") noted similar functions of paranormal experiences, with participants viewing experiences as a form of spiritual practice which helped them to confirm their pre-existing beliefs about the paranormal. Similarly, van Elk (2017) asked participants open-ended questions relating to their interest and belief in the paranormal and found that participants often attributed their interest to a specific live-event (e.g., the death of a parent) or experience (e.g., seeing a spirit, or witnessing an object levitate or move without an apparent physical cause). These findings support those presented by Clarke (1995), whose qualitative analysis found personal experiences to be the main reason for belief in the paranormal, followed by the experiences of others (e.g., friends and family) and the media.

In addition to maintaining paranormal beliefs, paranormal experiences have been attributed to both adaptive and maladaptive functions. Work by Mathijsen (2010; 2012) used semi-structured interviews to explore adolescent experiences of spiritualism (e.g., seeing or communicating with spirits or other supernatural entities), and found that these paranormal experiences were often accompanied by feelings of anxiety. Mathijsen (2012), however, notes that many of these experiences also served as an active coping mechanism to compensate for a need to understand death, a need to feel connected to the deceased, and to feel reassured when faced with the fear of death. Drinkwater et al. (2017) reported similar findings in their qualitative study. While participants often reported feelings of apprehension at the time of their paranormal experience, they viewed these experiences as important, fulfilling, and positive for personal growth.

Qualitative studies in this area have also explored the descriptions of paranormal sceptics to allow for some comparison of the importance of paranormal experiences for the maintenance of belief versus disbelief. Clarke (1995) noted that no personal paranormal experiences and insufficient evidence were the main reasons cited for disbelief. Schriever's (2000) qualitative analysis of sceptics' and believers' motives for (dis)belief, however, present a different account. Believers and sceptics were found to use the same motives, but in different frequencies. For example, both groups used personal experiences and friends to justify their (dis)belief, but believers used these motives significantly more than sceptics. Both groups used science and a lack of knowledge to justify their (dis)belief, but sceptics used these motives significantly more than believers. Both believers and sceptics, however, used motives such as rationality and communication to justify their (dis)belief in almost equal amounts. The findings from both Clarke (1995) and Schriever (2000) provide an interesting insight into the relationship between paranormal beliefs and experiences. Believers and sceptics both appear to base their belief (or lack thereof) on personal experience (Schriever, 2000), but where believers base their belief on specific experiences they have labelled as "paranormal", sceptics base their disbelief on the lack of an ostensibly paranormal experience (Clarke, 1995).

Despite the small volume of existing qualitative research, several key themes can be noted. First, that both paranormal beliefs and experiences have adaptive and maladaptive consequences. Second, that paranormal experiences are important for the maintenance of both paranormal beliefs and scepticism.

The present study

The present study builds upon the small foundation of qualitative research into belief in the paranormal, seeking to compare qualitative memory descriptions of believers and sceptics. Previous research highlighted in the Systematic Review chapter of this thesis provided inconsistent findings on the relationship between paranormal beliefs and memory. While some studies suggested that paranormal beliefs might be associated with an increased tendency to create false memories, Gray and Gallo's (2016) large-scale study suggested little quantitative support for differences in false memory generation, working memory updating ability, episodic or autobiographical memories. Qualitative differences, however, may exist in the subjective episodic memories of believers and sceptics, such that both groups demonstrate the same quantitative level of memory performance with qualitatively different content. The study, therefore, aimed to qualitatively explore the link between subjective paranormal experiences and belief (or *dis*belief) in paranormal phenomena through personal episodic memory descriptions of believers and sceptics. The study takes an exploratory approach, with no specific hypotheses made regarding the qualitative content of participants' responses.

STUDY 5.1. PARANORMAL BELIEFS AND EXPERIENCES: A QUALITATIVE STUDY

METHOD

Participants

An opportunistic sample of the general public was recruited through advertisements placed on social media. Participants were informed that they would be asked to provide written descriptions of memories or experiences that may have influenced their views about the paranormal. Eligibility criteria required participants to be over the age of 18 and fluent in English. Of the 200 participants recruited for the study, 43 (22%) provided complete responses. The final sample consisted of 16 males, 25 females, with 2 participants choosing not to report their gender. Participants were aged between 19-69 (M = 35.88, SD = 16.42), and most held an undergraduate degree or higher (70.50%). Of the final sample, 54% were aged between 19 and 29 (n = 23), 12% were aged between 30 and 39 (n = 5), 09% were aged between 40 and 49 (n = 4), 09% were aged between 50 and 59 (n = 4), and 16% were aged between 60 and 69 (n = 7).

Materials

Questionnaires

Paranormal beliefs were assessed using the Paranormal and Supernatural Beliefs Scale (PSBS; Dean et al., 2021).

Qualitative items

Participants were instructed to recall and type a description (in the box provided) of specific memories relating to paranormal beliefs and experiences. These four questions probed participants to reflect on: (1) a specific event that affected their belief in the paranormal, (2) whether their beliefs had changed over time (and the reason they attributed to this), (3) whether they had experienced or perceived something they had no rational explanation for (and if they thought this was a supernatural or paranormal phenomenon), and (4) whether they ever perceive things that other people cannot sense (and what this is like for them). Participants were asked to be as detailed as possible when providing their descriptions.

Procedure

The study was administered online using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020; www.gorilla.sc). All data were collected between 29 Oct 2021 and 30 Jan 2022. Participants were first provided with an information sheet detailing the aims of the study. Those providing their consent to participate in the study were then asked to provide their age, gender (male, female, other), and level of education (doctoral degree, postgraduate degree, undergraduate degree, post-secondary, secondary, vocational). Respondents were provided the option not to disclose these demographic details. The demographic section of the study also asked participants to self-identify themselves as a paranormal believer or sceptic, but also provided the option to state that they were unsure. Participants were then presented with the PSBS, followed by the four qualitative items.

Informed consent was obtained from all participants and all methods were performed in accordance with relevant guidelines and regulations. The study received ethical approval from the University of Hertfordshire Health, Science, Engineering and Technology Ethics Committee with Delegated Authority (HSET ECDA; protocol number LMS/PGR/UH/04716). The study was preregistered, and the protocol can be found on the Open Science Framework (OSF; https://osf.io/kha3v).

Data analysis

The corpus of memory descriptions was analysed using Alceste (IMAGE, 2018), a French software package for the statistical analysis of textual data. This software blends qualitative and quantitative methods to identify groups of words, phrases, and sentences that cluster together across different contexts. The resulting output presents categories of dominant themes (classes) for which the analyst assigns appropriate labels. ALCESTE has been widely used in many qualitative and mixed-methods studies (e.g., Valadez et al., 2020; Akhtar et al., 2018; Lelorain et al., 2012).

The present sample were divided into 'believers' and 'sceptics' according to their mean total scores on the PSBS. Participants scoring above the mean of 17.84 were classified as 'believers' (n = 28), while those scoring below were classified as sceptics (n = 15). Believers' and sceptics' memory descriptions were analysed separately to allow for comparisons between the two groups. Owing to the limited amount of data, believers' and sceptics' responses to all four items were analysed simultaneously (rather than individual analyses conducted for the groups' responses to each item). Labels for each class were assigned according to the qualitative responses (words, phrases, and sentences) that had clustered within them, and were corroborated with an additional researcher.

RESULTS

Table 5.1. presents a summary of participants' self-identification as a paranormal believer, sceptic, or unsure, in addition to the mean PSBS score for each group. Of the sample, most self-identified as a believer, and participants' self-categorisation was clearly reflected in the mean PSBS scores for each group. Self-identified believers demonstrated the highest mean PSBS scores (M = 24.47), followed by those who were unsure (M = 17.55), with self-identified sceptics scoring the lowest (M = 8.38).

Table 5.1. Frequencies of paranormal self-identification groups with mean and standard deviation PSBS scores.

Self-identification group	n	*M*(*SD*)
Believer	19	24.47 (5.52)
Sceptic	13	8.38 (6.79)
Unsure	11	17.55 (2.81)

Note: PSBS = Paranormal and Supernatural Beliefs Scale

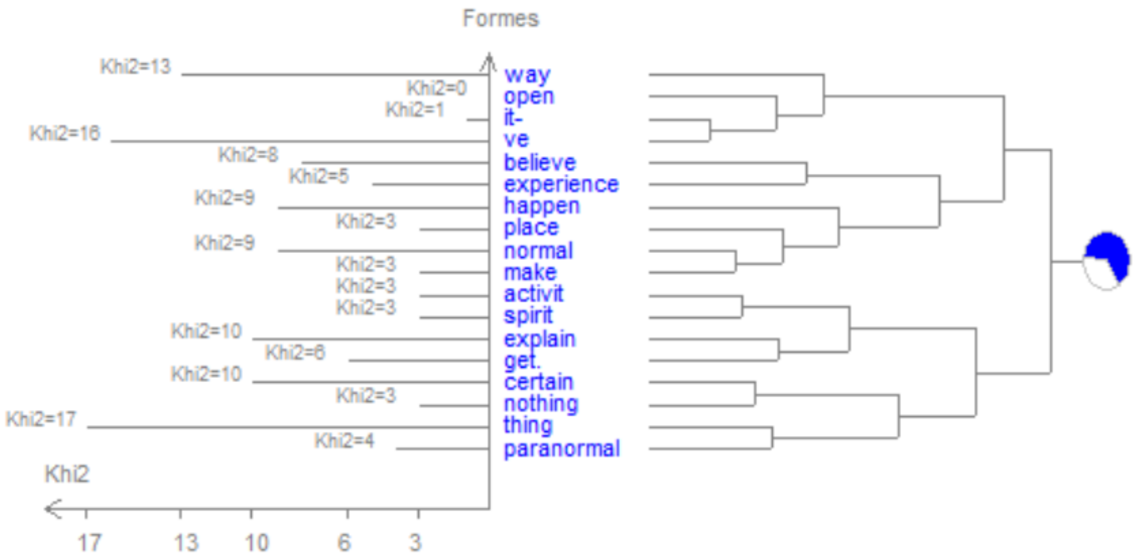
Believers

The analysis yielded five classes encompassing 52% of the descriptions provided by believers. This means that 48% of believers' memory descriptions could not be assigned to one of the identified classes. Table 5.2. presents a summary of each class with examples of participant responses. The dominant class (accounting for 24.64% of classified descriptions) concerned memory descriptions that were offering explanations of subjective paranormal experiences. Figure 5.1. depicts the words used most frequently in this class through hierarchical classification, including "I" (χ^2 = 28), "thing" (χ^2 = 17), and "explain" (χ^2 = 10). The hierarchical classification also shows associations between the identified terms. For example, "believe" was associated with "experience", "certain" with "nothing", and "activity" with "spirit" (see Figure 5.1. for a full representation of the associations between terms in this class). The grammatical category most strongly associated with this class were modal words ($\chi^2 = 3$; in descending order: “can”, “not”, “cannot”, “will”), while irregular verbs were most distant from this class ($\chi^2 = -4$; in descending order: “get”, “make”, “go”, “come”, “find”, “take”, “tell”, “dream”).

Table 5.2. Percentage of memories within each semantic category for paranormal believers in descending order.

Memory category	Percentage of memories	Example
Explaining experiences	24.64%	…many things which we as humans cannot explain scientifically… …only certain people have paranormal abilities… …science can explain a lot of things that I can't…
Intuitions	23.19%	…things I see or feel… …like having a 6th sense… …very intuitive yes…having a bad feeling about something and then something bad happens…
Perceptual experiences	18.84%	…seeing shadow figures… …hearing someone call my name…when no one else apart from myself is home… …felt her sitting next to me…
Spiritual experiences	18.84%	Lucid dreaming through a person that I don't know's eyes… …confirm this was indeed a prophecy… I know it was an act of God…
Explicit memories of experiences	14.49%	…an old steakhouse called the old homestead steakhouse…in a very old part of the city… …went up to the first floor where there were more tables… …went to Cornwall…rented a house on the coast…nearby lighthouse…

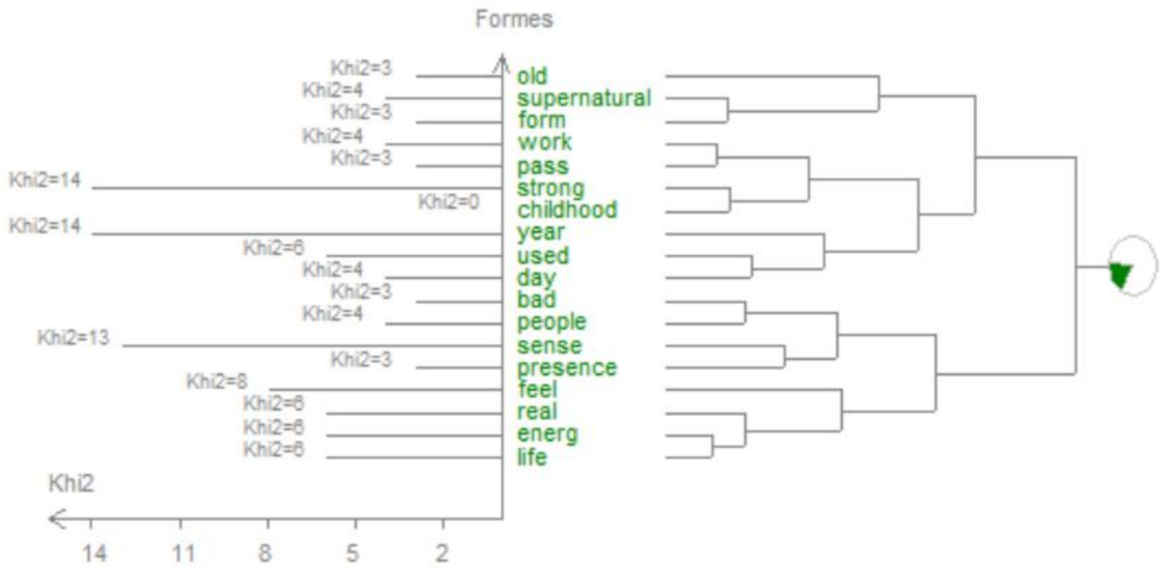
Figure 5.1. Dendrogram showing the hierarchical classification of words in the “Explaining experiences” class.



Note: Leftmost side of the figure depicts chi-square values for the most frequently occurring words in the class, with the righthand side showing the associations between words.

The next highest class (accounting for 23.19% of classified descriptions) concerned memory descriptions relating to intuitions. Figure 5.2. depicts the words used most frequently in this class through hierarchical classification, including “strong” ($\chi^2 = 14$), “sense” ($\chi^2 = 13$), and “feel” ($\chi^2 = 8$). The hierarchical classification identified associations between “sense” and “presence”, “energy” and “life”, and “supernatural” and “form” (see Figure 5.2. for a full representation of the associations between terms in this class). The grammatical category most strongly associated with this class were words indicating intensity ($\chi^2 = 4$; in descending order: “many”, “more”), while irregular verbs were most distant from this class ($\chi^2 = -3$; in descending order: “feel”, “find”, “come”, “go”, “say”, “see”, “hear”, “leave”, “get”).

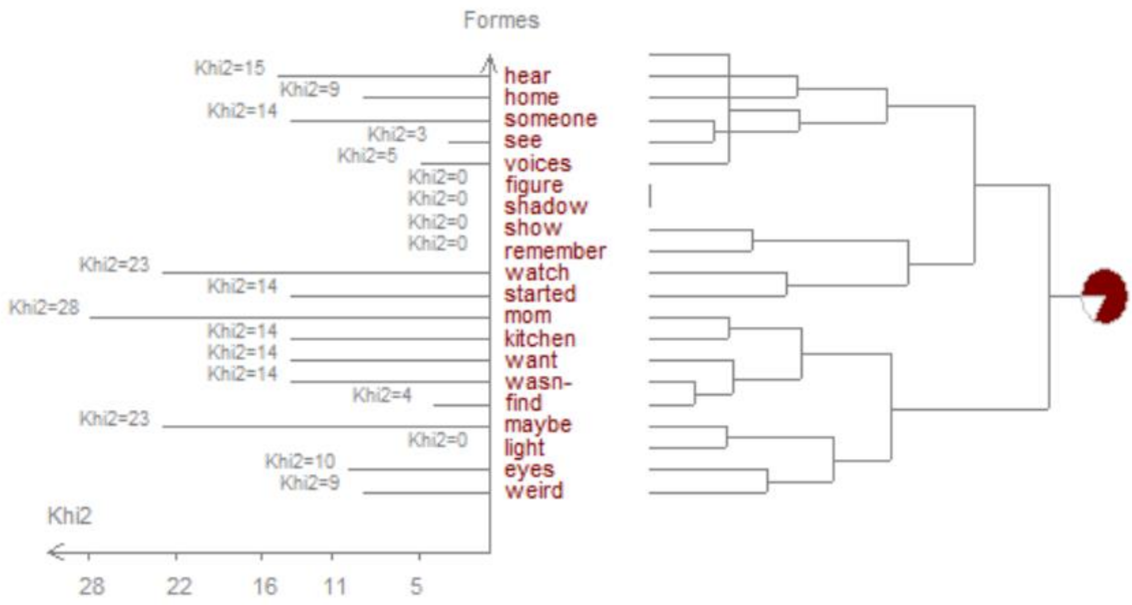
Figure 5.2. Dendrogram showing the hierarchical classification of words in the "Intuitions" class.



Note: Leftmost side of the figure depicts chi-square values for the most frequently occurring words in the class, with the righthand side showing the associations between words.

The next class (accounting for 18.84% of classified descriptions) concerned perceptual experiences of the paranormal. Figure 5.3. depicts the words used most frequently in this class, such as "watch" ($\chi^2 = 23$), "hear" ($\chi^2 = 15$), and "eyes" ($\chi^2 = 10$). The hierarchical classification identified associations between "someone" and "see", "watch" and "started", and "eyes" and "weird" (see Figure 5.3. for a full representation of the associations between terms in this class). The grammatical category most strongly associated with this class were irregular verbs ($\chi^2 = 4$; in descending order: "hear", "find", "see", "take", "think", "go", "make", "dream", "tell", "leave"), while words indicating intensity were most distant from this class ($\chi^2 = -4$; in descending order: "more").

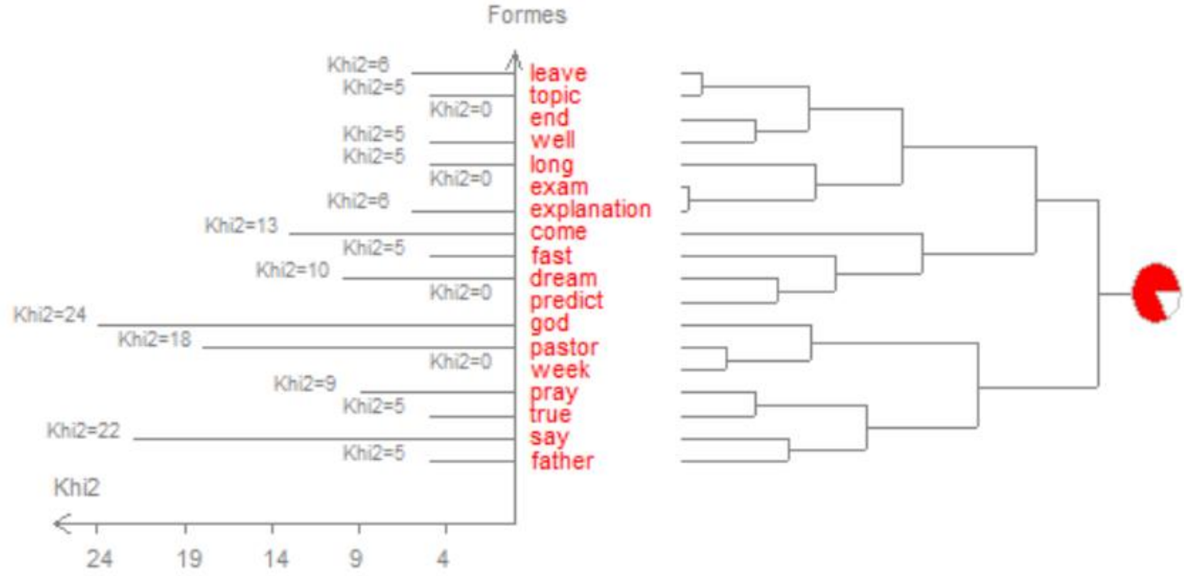
Figure 5.3. Dendrogram showing the hierarchical classification of words in the "Perceptual experiences" class.



Note: Leftmost side of the figure depicts chi-square values for the most frequently occurring words in the class, with the righthand side showing the associations between words.

The following class also accounted for 18.84% of classified descriptions and concerned spiritual experiences. Figure 5.4. depicts the words used most frequently in this class, such as "god" ($\chi^2 = 24$), "pastor" ($\chi^2 = 18$), and "dream" ($\chi^2 = 10$). The hierarchical classification identified associations between "dream" and "predict", "pray" and "true", and "end" and "well" (see Figure 5.4. for a full representation of the associations between terms in this class). The grammatical category most strongly associated with this class were irregular verbs ($\chi^2 = 2$; in descending order: "say", "come", "dream", "leave", "feel", "find", "make", "tell"), while prepositions indicating a movement were most distant from this class ($\chi^2 = -2$; in descending order: "up", "to").

Figure 5.4. Dendrogram showing the hierarchical classification of words in the "Spiritual experiences" class.



Note: Leftmost side of the figure depicts chi-square values for the most frequently occurring words in the class, with the righthand side showing the associations between words.

The final class accounted for 14.49% of classified descriptions and concerned explicit memories of experiences. Some of the most frequently used words in this class, such as "room" ($\chi^2 = 27$), "upstairs" ($\chi^2 = 19$), and "house" ($\chi^2 = 9$). The grammatical category most strongly associated with this class were prepositions indicating a movement ($\chi^2 = 6$; in descending order: "into"), while modal words were most distant from this class ($\chi^2 = -4$; in descending order: "no", "not"). Unfortunately, the hierarchical classification dendrogram identifying the frequency of, and association between, words in this class could not be produced during analyses (suggesting a lack of related concepts within this class).

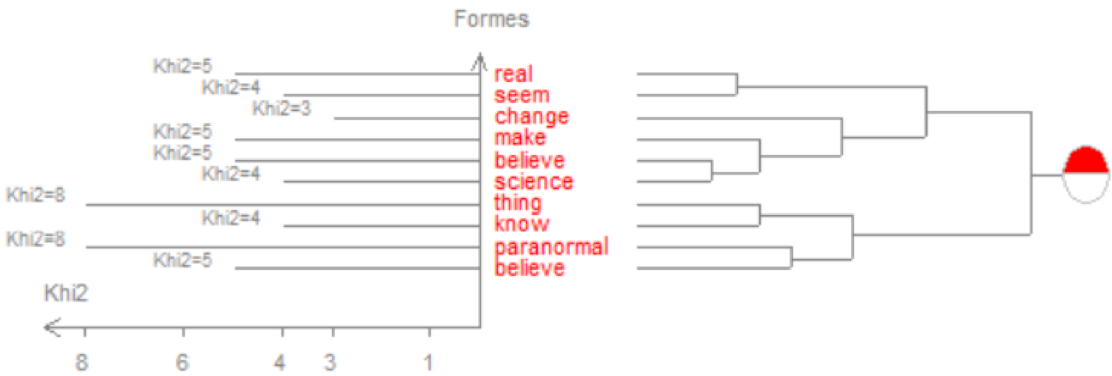
Sceptics

The analysis yielded two classes encompassing 93% of the descriptions provided by sceptics. This means that only 7% of sceptics' memory descriptions could not be assigned to one of the identified classes. Table 5.3. presents a summary of both classes with examples of participant responses. The dominant class (accounting for 62.50% of classified descriptions) concerned memory descriptions that were justifying disbelief. Figure 5.5. depicts the words used most frequently in this class through hierarchical classification, including “paranormal” ($\chi^2 = 8$), “real” ($\chi^2 = 5$), and “science” ($\chi^2 = 4$). The hierarchical classification identified associations between “believe” and “science”, “real” and “seem”, and “thing” and “know” (see Figure 5.5. for a full representation of the associations between terms in this class). The grammatical category most strongly associated with this class were words indicating intensity ($\chi^2 = 6$; in descending order: “many”, “everything”), while prepositions indicating a movement were most distant from this class ($\chi^2 = -12$; in descending order: “to”, “out”, “down”, “up”, “under”, “into”).

Table 5.3. Percentage of memories within each semantic category for paranormal sceptics in descending order.

Memory category	Percentage of memories	Example
Justifying disbelief	62.50%	...but it could be something else... ...discovered science and everything that seemed mystical could suddenly be explained away... ...just discovered science and everything I believed before stopped making sense...
Experiences with others	37.50%	...she and I saw something on the bed which looked...like a human... ...we heard a hum when we were waiting for someone... ...me and my father heard a response...

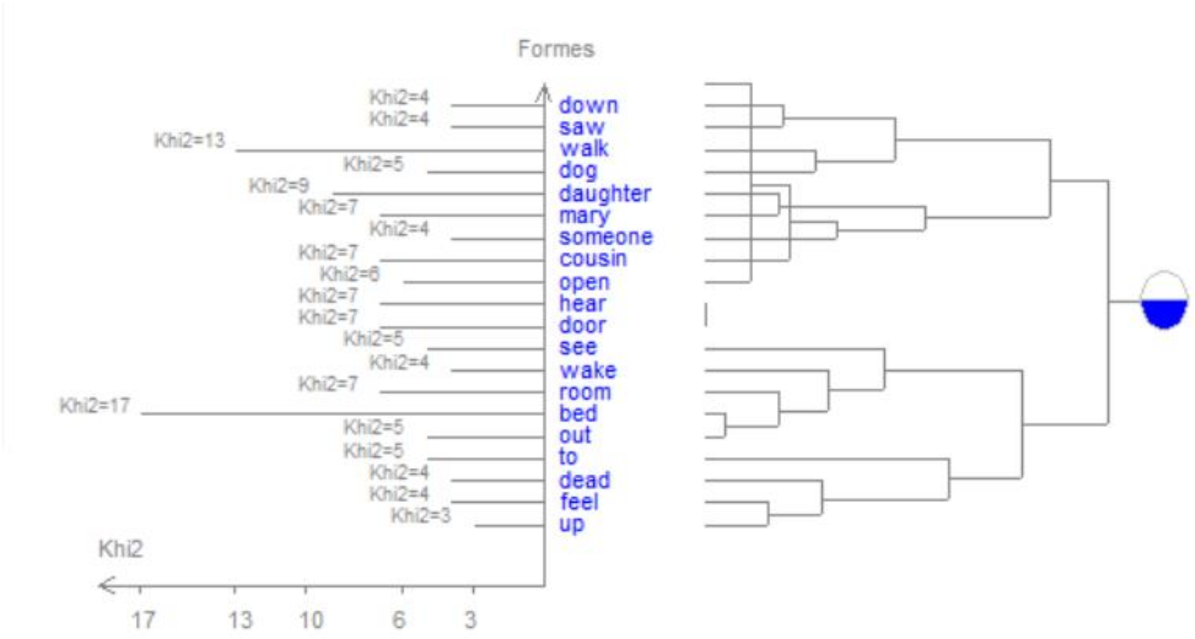
Figure 5.5. Dendrogram showing the hierarchical classification of words in the "Justification for disbelief" class.



Note: Leftmost side of the figure depicts chi-square values for the most frequently occurring words in the class, with the righthand side showing the associations between words.

The second class accounted for 37.50% of classified descriptions and concerned memory descriptions of experiences with others. Figure 5.6. depicts the words used most frequently in this class through hierarchical classification, including "daughter" ($\chi^2 = 9$), "her" ($\chi^2 = 7$), and "cousin" ($\chi^2 = 7$). The hierarchical classification identified associations between "walk" and "dog", "see" and "room", and "daughter" and "cousin" (see Figure 5.6. for a full representation of the associations between terms in this class). The grammatical category most strongly associated with this class were prepositions indicating a movement ($\chi^2 = 12$; in descending order: "to", "out", "down", "up", "under", "into"), while words indicating intensity were most distant from this class ($\chi^2 = -6$; in descending order: "many", "everything").

Figure 5.6. Dendrogram showing the hierarchical classification of words in the "Experiences with others" class.



Note: Leftmost side of the figure depicts chi-square values for the most frequently occurring words in the class, with the righthand side showing the associations between words.

Comparison of believers' and sceptics' memory descriptions

The total number of words provided by believers was 6473 (*M* = 1,618.25), while sceptics provided 3038 words (*M* = 759.50), suggesting that believers spent more time on their memory descriptions than sceptics. The experiences described by believers and sceptics shared similar themes but were expressed differently. For example, both groups mention hearing or seeing something unusual, such as voices they could not place or figures that disappeared (see Table 5.4. for examples of similar descriptions provided by believers and sceptics). However, where believers' descriptions were of personal experiences occurring directly to them, sceptics emphasised experiences that occurred with someone else present (see Table 5.5. for examples). Sceptics' memory descriptions were also more homogeneous than those provided by believers, as highlighted by the reduced number of classes identified and the higher percentage of classified data.

Table 5.4. Examples of similar memory descriptions provided by believers and sceptics.

Group	Example
Believer	…hearing voices which I cannot associate to anyone in the house… …seeing someone's face… …felt weight and warmth next to me on the couch…
Sceptic	…still sure that we heard a hum… …seeing a lady walk past the end of my bed… …felt as though someone of adult size was sat on the bed…

Table 5.5. Examples of experiences described by believers and sceptics.

Group	Example
Believer	…I was alone… …encountered my own experiences… …things I see or feel…
Sceptic	…witnessed by another person in the room with me… …experience was shared by daughter who was with me… …me and my father…

CHAPTER 5 GENERAL DISCUSSION

Paranormal experiences are, by their nature, subjective. The use of a qualitative study design here afforded an opportunity to explore the subjective experiences of a small sample of believers and sceptics, and how these experiences may have shaped their (dis)belief. For believers, 52% of memory descriptions were classified into five classes that express distinct components of their paranormal experiences: explaining experiences, intuitions, perceptual experiences, spiritual experiences, and semantic experiences. In contrast, 93% of sceptics' memory descriptions were classified into two classes: justification of disbelief, and experiences with others. Both believers and sceptics reported similar unusual experiences relating to the perception of something unexplained, such as hearing disembodied voices or seeing an unidentifiable figure. The groups differed, however, in the expression of these experiences, with believers emphasising personal experiences and sceptics emphasising experiences with others.

The present findings show stark similarity with those from Schriever (2000). Personal experiences were important for both belief and disbelief in the paranormal, with believers placing more emphasis on this than sceptics. The findings, therefore, do not strictly accord with Clarke's (1995) suggestion that sceptics' foundation for disbelief is a lack of personal experiences. The notion that sceptics report unusual or "paranormal" experiences is not strictly novel. For example, Dewan's (2013) interviews with sceptics in New Mexico (an area of growing paranormal interest since the alleged Roswell UFO crash in 1947) found several reported experiences of déjà vu, bizarre coincidences, and witnessing unusual aerial phenomena. Respondents felt, however, that their unusual experiences could either immediately be explained by science or could be if more information were available to them. These findings echo those reported by Lamont et al. (2009), who found that when faced with an ostensibly paranormal event for which no apparent explanation is plausible, sceptics appeal to an explanation that exists in principle but to which the details are unavailable to them (e.g., "…somehow there must be a way of doing that…", Lamont et al., 2009, pp. 550). In this vein, sceptics may ignore or dismiss inconvenient details in an effort to debunk an experience that cannot be accommodated within the boundaries of current scientific knowledge (Clarke, 2013). This suggestion could explain why believers' memory descriptions in the present study were more heterogeneous than those provided by sceptics. Where sceptics may have ignored or dismissed details of their experiences to conform with their scientific ideology, believers may have focused on specific details of their experiences (i.e., intuitive details, perceptual details, and semantic details) to conform with their own beliefs. Some support for this theory can be found in the quantitative results from Study 4.3. in which believers demonstrated weaker central coherence, with a bias towards detail-focused processing. Current models suggest that central coherence exists as a continuum in the general population in the form of a particular cognitive style, such that individuals with strong coherence process gist and global form at the expense of attention to (and memory for) detail, while those with weak coherence demonstrate good memory for details and verbatim information (Happé & Frith, 2006). It could be argued, using this conceptualisation, that paranormal believers lie at the detail-focused end of the normal continuum (while sceptics lie at the opposite global end) and therefore recall more specific details of events related to their belief in, and experience of, paranormal phenomena.

While Study 5.1. provides a useful perspective on the subjective importance of paranormal experiences for the development of (dis)belief, it is not without its limitations. Firstly, the high attrition rate of the study presents concerns for nonresponse bias. Nonresponse bias occurs when participants who fail to complete a study are systematically different from those who do complete the study, and this bias is thought to become a critical issue when response rates fall below 70% (Prince, 2012). Of the 200 participants recruited for the present work, only 43 (22%) completed the study. This raises questions about the reason for such a high attrition rate of 78%. As Betsch et al. (2021) note, those with paranormal beliefs often cite social exclusion as one of the biggest challenges to their beliefs. It may therefore be the case that fewer individuals completed the study for fear of social exclusion or negative judgement from others. Paranormal experiences (Betsch et al., 2021) and social exclusion (Graeupner & Coman, 2017), however, are thought to strengthen endorsement of paranormal beliefs, and so those completing the study by providing details of their personal paranormal experiences may hold stronger beliefs than those who failed to complete the study. Nonetheless, these accounts are more applicable to the response rates of paranormal believers than sceptics. As previously mentioned in this thesis, paranormal *dis*belief is a belief type in its own right (Lamont et al., 2009), and so sceptical participants may have failed to complete the study for similar reasons. While social exclusion is less pertinent to paranormal disbelief (as it conforms to the scientific-rationalistic orientation of social convention), sceptics with personal paranormal experiences may also fear social exclusion or negative perception from reporting these unusual experiences. Similarly, these personal experiences may strengthen endorsement of paranormal disbelief in much the same way as for paranormal beliefs, which could again suggest that sceptics completing the study held stronger *dis*belief than sceptics who failed to complete the study. This would accord with previous research suggesting both sceptics and believers base their (dis)belief on personal experiences (Schriever, 2000). The importance of personal experiences for the development of paranormal (dis)belief could therefore be overestimated if those with stronger (dis)beliefs are more likely to take part in studies such as this.

An additional limitation of the sample in this study was that it was biased towards believers (65%). The difference in the volume of words provided by believers and sceptics was an interesting finding from the study. It is possible, however, that this is simply an effect of the larger number of believers compared to sceptics ($n = 28$ vs. $n = 15$). This may also account for some of the heterogeneity noted in believers' responses. Finally, while Alceste reduces (although does not eliminate) researcher bias, the software emphasises collective representations of the corpus of memory descriptions rather than providing an analysis of important minor insights. Use of this software does not prevent complementary qualitative analysis of these minor insights; however, this would increase researcher bias for any findings.

Study 5.1. demonstrates the value of considering the subjective importance of paranormal experiences for both believers and sceptics to better comprehend the nature of paranormal (dis)belief. It is suggested that future qualitative research does not limit data collection to believers (as is often common practice). Knowing how believers *and* sceptics experience unusual or "paranormal" events provides a tool that could serve to better understand the development and maintenance of belief versus disbelief in paranormal phenomena. Further research may also benefit from considering potential nonresponse biases in relation to the subjective reporting of personal paranormal experiences.

CHAPTER 6. CONCLUSION

Review of results

This thesis took several approaches to assess the relationship between paranormal beliefs and cognitive functioning. The work first establishes a reliable and valid measure of paranormal beliefs in the general population, before considering the possible link between paranormal beliefs and deficits of cognitive functioning (specifically relating to executive functions). Finally, the work examines the role of personal experiences for the development of paranormal beliefs and scepticism.

With regards to the accurate measurement of paranormal beliefs in the general population, Chapter 3 demonstrated that the seven-factor structure of the widely used Revised Paranormal Belief Scale (RPBS; Tobacyk, 2004) was not the most prudent description of the scale, finding a reduced four-factor structure to be more appropriate. Despite this more appropriate factor structure, issues surrounding the item content and statistical techniques used to develop the scale indicated a need for a more refined and accurate measure of paranormal beliefs in the general population. While the classical test theory method of exploratory factor analysis was able to reduce an original corpus of 29 items to a 14-item single factor scale, it was unable to provide the same level of detailed assessment as that of the modern test theory Rasch method which produced the final 13-item Paranormal and Supernatural Beliefs Scale (PSBS; Dean et al., 2021). Specifically, the Rasch method allowed for an assessment of the functionality of the response scale (reducing this from a 7-point Likert scale to a 4-point scale), item response difficulty, and differential item functioning, to produce a reliable and valid measure of paranormal beliefs in the general population.

Future assessment of the scale's construct validity, however, is warranted. In particular, the convergent, divergent, and nomological validity of the PSBS was not assessed in the present work. The convergent validity of a scale is demonstrated by examining the extent to which measures of the same variable are correlated. A scale can, therefore, be considered to accurately represent the latent variable of interest if it correlates highly with other measures of the same variable (e.g., the ASGS or RPBS). In relation to divergent validity, concerned with establishing a measure's relative uniqueness, a scale should not correlate strongly with measures of similar but distinct traits with which it theoretically shares little common variance. By obtaining divergent validity, the latent variable measured by the PSBS (i.e., paranormal beliefs) can be considered as conceptually independent from other similar concepts. For example, Irwin and Marks (2013) argue that paranormal beliefs, typically associated with an intuitive thinking style, are generated and maintained by different psychological processes compared to other belief types that are derived from more analytical thinking. If the PSBS, therefore, failed to correlate well with a measure of more analytically founded beliefs (e.g., the Scientific Epistemic Beliefs scale; Lindfors et al., 2019), this would provide some indication of its divergent validity. Finally, nomological validity of the scale assesses the degree to which predictions from a theoretical network (e.g., correlational relationships) are confirmed (Holton III et al., 2007). Good nomological validity indicates that the statistical structure of the PSBS is aligned with the theoretical claims of paranormal beliefs within the literature. Considering the previous literature presented in the Systematic Review chapter of this thesis, and the idea that paranormal beliefs are related to a more intuitive thinking style rather than an analytical one, the PSBS could be considered to have good nomological validity if it demonstrates strong positive correlations with measures of intuitive thinking and strong negative correlations with measures of analytical thinking. For example, positive correlations should be seen with the Experientiality subscale of the Rational-Experiential Inventory (REI; Pacini & Epstein, 1999) and strong negative correlations with the Rationality subscale.

Chapter 4 presented a detailed assessment of the relationship between paranormal beliefs and executive functioning, with a particular focus on cognitive flexibility. The first study in this chapter found a negative relationship between paranormal beliefs and cognitive flexibility, supported by both frequentist and Bayesian statistics. A negative relationship was noted between paranormal beliefs and working memory updating ability, but Bayesian analyses indicated that only anecdotal evidence was present for this relationship. As the strongest relationship was between paranormal beliefs and cognitive flexibility, the remaining four studies in Chapter 4 sought to replicate the negative association. Contrary to the hypothesised relationship, Study 4.2. found a positive association between paranormal beliefs and cognitive flexibility, such that believers demonstrated increased cognitive flexibility compared to sceptics. It was initially thought that this finding offered a more accurate representation of the relationship between paranormal beliefs and cognitive flexibility owing to the suitably powered sample, however efforts to replicate this finding were not successful. Studies 4.3. and 4.4. supported a negative association between paranormal beliefs and cognitive flexibility, with sceptics demonstrating faster reaction times and fewer errors on the Wisconsin Card Sorting Test (WCST; see Berg, 1948; Grant &

Berg, 1948) compared to believers. It should be noted, however, that no significant association between paranormal beliefs and cognitive flexibility was found when examining only the WCST performance of participants with higher levels of education. Despite findings suggesting a negative association between paranormal beliefs and cognitive flexibility, meta-analysis combining data from Studies 4.2. to 4.4. found no significant difference in cognitive flexibility between believers and sceptics, with a small, pooled effect size and high heterogeneity. While these studies provided some indication that believers and sceptics differ in their cognitive flexibility, the findings are not consistent and suggest that any relationship existing between paranormal beliefs and cognitive flexibility is weak and likely mediated by an unknown additional variable.

The inconsistency of findings demonstrated throughout this chapter highlight the need for replication studies in this field. These replication studies should not only relate to executive functions, but also to cognitive functions more broadly. As highlighted in the Systematic Review chapter, similar inconsistencies can be seen when looking at different domains of cognitive function in relation to paranormal beliefs (e.g., perceptual decision-making and the conjunction fallacy). Studies seeking to determine the relationship between paranormal beliefs and cognitive function should, therefore, conduct replications with the goal of conducting meta-analytic assessments of the strength, and potential moderating factors, of the relationship. This would also help to lessen the effect of any publication bias existing in the field. As previously discussed, such bias might relate to the fact that studies are more likely to be published when they present significant of favourable findings (Song et al., 2009), and are generally rated as being of better quality when they conform to prior expectations or theoretical models. This could be reflective of the "file-drawer" effect, defined by Rosenthal (1979) in his seminal work as a tendency for researchers to publish studies that "work" and offer significant results, and place conflicting or non-significant studies in the file-drawer. Subsequently, published evidence might be unrepresentative of the actual effect or relationship between to variables. However, as suggested in the Systematic Review chapter, any replication studies should be preregistered with clear descriptions of the key outcome measures and analysis procedures. Not only will this allow for clearer interpretation of future study findings in an already inconsistent area of research, but this will also help to lessen the extent of publication biases, particularly relating to the file-drawer effect.

Finally, Chapter 5 used a qualitative approach to consider the subjective importance of personal paranormal experiences for the development and maintenance of paranormal (dis)belief. Believers overall provided more detailed memory descriptions compared to sceptics and focused on personal paranormal experiences rather than the shared experiences reported by sceptics. The memory descriptions provided by believers could be classified into five distinct classes representing components of their paranormal experiences: explaining experiences, intuitions, perceptual experiences, spiritual experiences, and semantic experiences. Sceptics' memory descriptions were classified into two distinct classes: justification of disbelief, and experiences with others. Sceptics' descriptions presented a more homogeneous account of their paranormal experiences and the subjective effect these had on their disbelief. The more detailed memory descriptions provided by paranormal believers provided some support for the quantitative findings in Study 4.3., which suggested believers have a weaker central coherence and a bias towards detail-focused processing. Using current central coherence models (e.g., Happé & Frith, 2006), it could be argued that paranormal believers lie at the detail-focused end of a normal continuum, and therefore recall more specific details of events, compared to sceptics who lie at the opposite global-processing end of the continuum. This theory warrants additional testing in future research, however, as the biased sample size in the present study (65% believers) might have contributed to the larger volume of words used in responses from believers compared to sceptics.

Limitations and methodological issues

One limitation of the present work, particularly relevant for Chapter 2, is that differing measures of paranormal beliefs have been used. While most of the work presented in this thesis makes use of the Rasch-developed PSBS (Dean et al., 2021), two of the five studies in Chapter 2 used different measures (Tobacyk's, 2004, RPBS and the factor analytic 14-item version of the PSBS, Dean et al., 2021). This may make it difficult to generalise across these studies about the cognitive functioning of believers and sceptics. It should be noted, however, that the 14-item PSBS (Dean et al., 2021) demonstrates a strong positive correlation with the final version of the scale (see Dean et al., 2021).

Samples in each of the studies across the chapters presented here were largely biased towards white, well-educated, female participants. It can be argued, therefore, that these samples are not necessarily representative of the general population. The issue of gender is important to consider in relation to several of the studies presented here. While differences introduced by gender are not strictly relevant for paranormal beliefs where scales free of differential item functioning have been used (see Lange et al., 2000), they are pertinent to consider when examining cognitive functioning. Some studies have noted gender differences in cognitive flexibility (Wang et al., 2022; Esen-Aygun, 2018), though these differences are often small and have been recorded in opposing directions. While most studies note no significant gender differences on either subjective or objective measures of cognitive flexibility (Zmigrod et al., 2019; Kercood et al., 2017; Douw et al., 2016; Johnco et al., 2015), the few studies reporting marginal influences of gender on cognitive flexibility make consideration of gender as potential confounding variable advisable. There is also evidence to suggest gender differences in the recall of personal memories. Females have been found to produce more detailed and vivid memory descriptions than males (Ross & Holmberg, 1990), showing a preference for episodic memory descriptions (Pillemer et al., 2003). This introduces a potential confounding variable when examining the volume and content of memory descriptions provided by believers and sceptics in the final study presented in this thesis. For example, while the number of males was almost equal in both groups (five sceptics and six believers), there were more females in the 'believer' group ($n = 9$) compared to the 'sceptic' group ($n = 3$). This gender bias may have contributed to the difference noted in the volume of words produced by believers and sceptics in Study 5.1.

Directions for future research

The central argument of the cognitive deficits hypothesis (see Irwin, 1993) is that paranormal believers show marked deficits in cognitive functioning compared to sceptics. Some support for this relationship was found through the replication studies conducted in Chapter 4, with believers demonstrating reduced cognitive flexibility compared to sceptics in three of the five studies. If paranormal beliefs were characterised by a global cognitive deficit, however, then believers should have demonstrated consistently reduced performance across all cognitive measures used in the studies presented in Chapter 4 (including each behavioural measure of cognitive flexibility, and measures of inhibitory control and working memory updating ability). This trend should also be seen in the wider literature but as noted in the systematic literature review (Chapter 2), the findings relating to cognitive deficits and paranormal beliefs are not homogenous. The lack of consensus surrounding cognitive deficits and paranormal beliefs suggests that the cognitive deficits hypothesis is not an accurate depiction of the relationship between cognition and belief in the paranormal. This presents the need for a new theory, such as the fluid-executive theory (Dean et al., 2022, see Chapter 2 of this thesis), which argues that cognitive *differences* between believers and sceptics may be accounted for by their levels of fluid intelligence and higher-order executive functioning. It would be interesting to test this theory by grouping believers and sceptics into those with higher or lower fluid intelligence and comparing potential main and interaction effects of group membership on a range of cognitive tasks. For example, using this method to test for cognitive differences between believers and sceptics in the areas identified by the systematic literature review as having the highest heterogeneity (e.g., memory, critical thinking ability, perceptual decision-making) would help to clarify the relationship between paranormal beliefs and these aspects of cognition. It would also be useful to use this method to re-examine the relationship between paranormal beliefs and cognitive flexibility to confirm and clarify the reduced cognitive flexibility identified for believers in Chapter 4.

In addition to quantitatively testing the fluid-executive theory (Dean et al., 2022), future research should also seek to qualitatively (and quantitatively) explore sceptics' experiences with the paranormal. Research in this area almost exclusively focuses on paranormal beliefs, effectively using scepticism, or *dis*belief, as the criterion for a control group. In this sense, research has traditionally approached belief in the paranormal as problematic in ways that disbelief is not (Lamont et al., 2009). As scepticism can be considered as a belief type in its own right (Lamont et al., 2009), it is worthy of as much study as that devoted to paranormal beliefs. Not only should future research consider the cognitive profiles of both belief *and* disbelief, but also the subjective benefits of these belief types. Previous work has suggested that paranormal beliefs act as a cognitive 'defence' against the uncertainty of life events (Williams & Irwin, 1991), with qualitative research noting that these beliefs foster positive feelings of resilience, hope, and coping following a bereavement (Cox et al., 2017). It would therefore be interesting to conduct qualitative research into the subjective benefits of scepticism, and the impact individuals feel their disbelief has for daily living and significant life events. Finally, future work should consider testing central coherence models (e.g., Happé & Frith, 2006) in relation to paranormal beliefs to explore the suggestion made in Chapter 5 that paranormal believers demonstrate a cognitive bias towards detail-focused processing affecting their memory for specific details and verbatim information.

Concluding remarks

The aims of this thesis were threefold. Firstly, to establish a reliable and valid measure of paranormal beliefs in the general population. Secondly, to consider the possible link between paranormal beliefs and deficits of cognitive functioning, with a specific focus on executive functions. Finally, to examine the role of personal experiences for the development of paranormal beliefs and scepticism.

The work concerning the reliable measurement of paranormal beliefs in the general population noted several conceptual and statistical problems with the widely used RPBS (Tobacyk, 2004), leading to the development of a more reliable and accurate measure. Two methods for scale development (the classical test theory method of factor analysis, and the modern test theory method of Rasch analysis) were employed to reduce an initial corpus of 29 items to the most prudent measure of paranormal beliefs in the general population. The final Rasch-developed measure was labelled the Paranormal and Supernatural Beliefs Scale (PSBS; Dean et al., 2021), and uses a 4-point Likert scale to assess participants' endorsement of 13 items relating to various phenomena. When considering the second aim of this work, there was some indication that sceptics may be more cognitively flexible than believers. The replication of this finding, however, presented many challenges. Of five studies, three suggested a negative association between paranormal beliefs and cognitive flexibility, with one finding a positive association, and one finding no significant association. Combining the data collected in each of these studies revealed no significant difference between believers' and sceptics' cognitive flexibility, with a small effect and high heterogeneity across the studies. Finally, work concerning the third aim of this thesis highlighted several differences in the subjective paranormal experiences of believers and sceptics and how these experiences contributed to (dis)belief. The memory descriptions provided by believers were longer and more heterogeneous than those provided by sceptics. Believers also emphasised personal experiences, whereas sceptics emphasised experiences with others present.

In addition to the development of a reliable measure of paranormal beliefs in the general population, and the qualitative exploration of the subjective importance of paranormal experiences for the development and maintenance of (dis)belief, this thesis sought to determine whether belief in the paranormal is associated with cognitive deficits. From the evidence presented here, it is apparent that while there may be cognitive *differences* between paranormal believers and sceptics, this does not equate to a global cognitive deficit.

REFERENCES

Aarnio, K., & Lindeman, M. (2005). Paranormal beliefs, education, and thinking styles. *Personality and Individual Differences, 39*(7), 1227-1236.

Aarnio, K., & Lindeman, M. (2007). Religious people and paranormal believers: Alike or different? *Journal of Individual Differences, 28*(1), 1-9.

Agnew, J.H., & Bidwell, W.H. (1863). Illusions and Hallucination. *Eclectic Magazine: Foreign Literature, 58*, 41-51.

Akhtar, S., Justice, L.V., Morrison, C.M., & Conway, M.A. (2018). Fictional first memories. *Psychological Science, 29*(10), 1612-1619.

Alba, A.C., Alexander, P.E., Chang, J., Maclsaac, J., DeFry, S., & Guyatt, G.H. (2016). High statistical heterogeneity is more frequent in meta-analysis of continuous than binary outcomes. *Journal of Epidemiology, 70*, 129-135.

Alcock, J.E., & Otis, L.P. (1980). Critical thinking and belief in the paranormal. *Psychological Reports, 46*, 479-482.

Amso, D., Haas, S., McShane, L., & Badre, D. (2014). Working memory updating and the development of rule-guided behaviour. *Cognition, 133*(1), 201-210.

Andrews, R.A., & Tyson, P. (2019). The superstitious scholar: Paranormal belief within a student population and its relationship to academic ability and discipline. *Journal of Applied Research in Higher Education, 11*(3), 415-427.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Anwyl-Irvine, A.L., Massonié, J., Flitton, A., Kirkham, N.Z., & Evershed, J.K. (2020). Gorilla in our midst: An online behavioural experiment builder. *Behaviour Research Methods, 52*(1), 388-407.

Bader, C.D., Baker, J.O., & Molle, A. (2012). Countervailing forces: Religiosity and paranormal belief in Italy. *Journal for the Scientific Study of Religion, 51*(4), 705-720.

Baggetta, P., & Alexander, P.A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education, 10*(1), 10-33.

Baker, J.O., & Draper, S. (2010). Diverse supernatural portfolios: Certitude, exclusivity, and the curvilinear relationship between religiosity and paranormal beliefs. *Journal for the Scientific Study of Religion, 49*(3), 413-424.

Ballová Mikušková, E., & Čavojová, V. (2020). The effect of analytic cognitive style on credulity. *Frontiers in Psychology, 11*, 2682.

Banich, M.T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18*(2), 89-94.

Barberia, I., Tubau, E., Matute, H., & Rodríguez-Ferreiro, J. (2018). A short educational intervention diminishes causal illusions and specific paranormal beliefs in undergraduates. *PLoS One, 13*(1), e0191907.

Basso, M.R., Bornstein, R.A., & Lang, J.M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist, 13*(3), 283-292.

Beck, R., & Miller, J.P. (2001). Erosion of belief and disbelief: Effects of religiosity and negative affect on beliefs in the paranormal and supernatural. *The Journal of Social Psychology, 141*(2), 277-287.

Berg, E.A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology, 39*, 15-22.

Berkowski, M., & MacDonald, D.A. (2014). Childhood trauma and the development of paranormal beliefs. *The Journal of Nervous and Mental Disease, 202*(4), 305-312.

Betsch, T., Aßmann, L., & Glöckner, A. (2020). Paranormal beliefs and individual differences: Story seeking without reasoned review. *Heliyon, 6*(6), e04259.

Betsch, T., Jäckel, P., Hammes, M., & Brinkmann, B.J. (2021). On the adaptive value of paranormal beliefs – a qualitative study. *Integrative Psychological and Behavioural Science*, 1-11.

Blackmore, S.J. (1997). Probability misjudgement and belief in the paranormal: A newspaper survey. *British Journal of Psychology, 88*(4), 683-689.

Blackmore, S., & Moore, R. (1994). Seeing things: Visual recognition and belief in the paranormal. *European Journal of Parapsychology, 10*, 91-103.

Blackmore, S., & Trośianko, T. (1985). Belief in the paranormal: Probability judgements, illusory control, and the 'chance baseline shift'. *British Journal of Psychology, 76*(4), 459-468.

Blanco, F., Barberia, I., & Matute, H. (2015). Individuals who believe in the paranormal expose themselves to biased information and develop more causal illusions than nonbelievers in the laboratory. *PloS One, 10*(7), e0131378.

BMG Research. (2017). *BMG Halloween Poll: A third of Brits believe in ghosts, spirits, or other types of paranormal activity.* BMG Research. https://www.bmgresearch.co.uk/bmg-halloween-poll-third-brits-believe-ghosts-spirits-types-paranormal-activity/

Bösch, H., Steinkamp, F., & Boller, E. (2006). Examining psychokinesis: The interaction of human intention with random number generators – A meta analysis. *Psychological Bulletin, 132*(4), 497-523.

Böttinger, E. (1976). Einstellung zum 'Okkulten' und Personlichkeitsmerkmale [Attitude to the 'occult' and personality traits]. *Zeitschrift fur Parapsychologie und Grenzgebiete der Psychologie, 1*, 116-128.

Bouvet, R., Djeriouat, H., Goutaudier, N., Py, J., Chabrol, H. (2014). Validation française de la Revised Paranormal Belief Scale [French validation of the Revised Paranormal Belief Scale]. *L'Encéphale, 40*(4), 308-314.

Branković, M. (2019). Who believes in ESP: Cognitive and motivational determinants of the belief in extra-sensory perception. *Europe's Journal of Psychology, 15*(1), 120-139.

Bressan, P. (2002). The connection between random sequences, everyday coincidences, and belief in the paranormal. *Applied Cognitive Psychology, 16*(1), 17-34.

Brewer, P.R. (2012). The trappings of science: Media messages, scientific authority, and beliefs about paranormal investigators. *Science Communication, 35*(3), 311-333.

Broad, C.D. (1949). The relevance of psychical research to philosophy. *Philosophy, 24*(91), 291-309.

Brotherton, R., & French, C.C. (2014). Belief in conspiracy theories and susceptibility to the conjunction fallacy. *Applied Cognitive Psychology, 28*(2), 238-248.

Brugger, P., & Baumann, A.T. (1994). Repetition avoidance in responses to imaginary questions: The effect of respondents' belief in ESP. *Psychological Reports, 75*, 883-893.

Brugger, P., Landis, T., & Regard, M. (1990). A 'sheep-goat effect' in repetition avoidance: Extra-sensory perception as an effect of subjective probability? *British Journal of Psychology, 81*(4), 455-468.

Brugger, P., Regard, M., & Landis, T. (1991). Belief in extrasensory perception and illusory control: A replication. *Journal of Psychology, 125*(4), 501-502.

Buhr, K., & Dugas, M.J. (2002). The intolerance of uncertainty scale: psychometric properties of the English version. *Behaviour Research and Therapy, 40*, 931-945.

Burger, J.M., & Lynn, A.L. (2005). Superstitious behaviour among American and Japanese professional baseball players. *Basic and Applied Social Psychology, 27*(1), 71-76.

Burgess, P.W., Alderman, N., Wilson, B.A., Evans, J.J., & Emslie, H. (1996). The dysexecutive questionnaire. *Behavioural Assessment of the Dysexecutive Syndrome*. Bury St. Edmunds, UK: Thames Valley Test Company.

Burgess, P.W., & Shallice, T. (1997). *The Hayling and Brixton Tests*. Thurston, England: Thames Valley Test Company.

Burnham, K.P., & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). Springer-Verlag New York, Inc.

Bryan, C.J., Tipton, E., & Yeager, D.S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*(8), 980-989.

Callaghan, A.J., & Irwin, H.J. (2003). Paranormal belief as a psychological coping mechanism. *Journal of the Society for Psychical Research, 67*(872), 200-207.

Capa, R.L., Bouquet, C.A., Dreher, J.C., & Dufour, A. (2013). Long-lasting effects of performance-contingent unconscious and conscious reward incentives during cued task-switching. *Cortex, 49*(7), 1943-1954.

Caputo, G.B. (2015). Dissociation and hallucinations in dyads engaged through interpersonal gazing. *Psychiatry Research, 228*(3), 659-663.

Caputo, G.B. (2017). Strange-face illusions during interpersonal-gazing and personality differences of spirituality. *Explore, 13*(6), 379-385.

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist, 73*(5), 663-677.

Cattell, R.B. (1963). Theory of fluid and crystallised intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22.

Chandler, J., Mueller, P., & Paolacci, G. (2013). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioural researchers. *Behavioural Research Methods, 46*, 112-130.

Claridge, G. E. (1997). *Schizotypy: Implications for illness and health.* Oxford University Press.

Claridge, G., & Beech, T. (1995). Fully and quasi-dimensional constructions of schizotypy. In A. Raine, T. Lencz, & S.A. Mednick (Eds.), *Schizotypal Personality* (pp. 192-216). Cambridge University Press.

Clark, M.S., Dennerstein, L., Elkadi, S., Guthrie, J.R., Bowden, S.C., & Henderson, V.W. (2004). Normative data for tasks of executive function and working memory for Australian-born women aged 56-67. *Australian Psychologist, 39*(3), 244-250.

Clarke, D. (1991). Belief in the paranormal: A New Zealand survey. *Journal of the Society for Psychical Research, 57*(823), 412-425.

Clarke, D. (1995). Experience and other reasons given for belief and disbelief in paranormal and religious phenomena. *Journal of the Society for Psychical Research, 60*(841), 371-384.

Clarke, D. (2013). Extraordinary experiences with UFOs. In O. Jenzen, & S.R. Munt (Eds.), *The Ashgate Research Companion to Paranormal Cultures* (pp. 79-93). Routledge.

Clarke, E., & Kiropoulos, L.A. (2021). Mediating the relationship between neuroticism and depressive, anxiety and eating disorder symptoms: The role of intolerance of uncertainty and cognitive flexibility. *Journal of Affective Disorders Reports, 4*, 100101.

Cohen, J. (1988). *Statistical power analysis for the social sciences* (2nd Ed.). Lawrence Erlbaum Associates.

Cohen, A.S., Matthews, R.A., Najolia, G.M., & Brown, L.A. (2010). Toward a more psychometrically sound brief measure of schizotypal traits: Introducing the SPQ-Brief Revised. *Journal of Personality Disorders, 24*(4), 516-537.

Colzato, L.S., van Wouwe, N.C., Lavender, T.J., & Hommel, B. (2006). Intelligence and cognitive flexibility: Fluid intelligence correlated with feature "unbinding" across perception and action. *Psychonomic Bulletin & Review, 13*(6), 1043-1048.

Comrey, A.L., & Lee, H.B. (2013). *A first course in factor analysis* (2nd ed.). Psychology Press.

Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Cox, C.J., Cooper, C.E., & Smith, M.D. (2017). Exploring the effects of mediumship on hope, resilience, and post-traumatic growth in the bereaved. *Journal of Exceptional Experiences and Psychology, 5*(2), 6-15.

Crawford, J.R., & Henry, J.D. (2005). Assessment of executive dysfunction. In P.W. Halligan & D.T. Wade (Eds.), *Effectiveness of rehabilitation for cognitive deficits* (pp. 233-245). Oxford University Press.

Dag, I. (1999). The relationships among paranormal beliefs, locus of control and psychopathology in a Turkish college sample. *Personality and Individual Differences, 26*(4), 723-737.

Dagnall, N., Denovan, A., Drinkwater, K., Parker, A., & Clough, P. (2016a). Toward a better understanding of the relationship between belief in the paranormal and statistical bias: The potential role of schizotypy. *Frontiers in Psychology, 7*, 1045.

Dagnall, N., Denovan, A., Drinkwater, K., Parker, A., & Clough, P.J. (2017). Urban legends and paranormal beliefs: The role of reality testing and schizotypy. *Frontiers in Psychology, 8*, 942.

Dagnall, N., Parker, A., Munley, G., & Drinkwater, K. (2010a). Common paranormal belief dimensions. *Journal of Scientific Exploration, 24*(3), 431-447.

Dagnall, N., Parker, A., Munley, G., & Drinkwater, K. (2010b). The relationship between belief in extra-terrestrial life, UFO-related beliefs and paranormal belief. *Journal of the Society for Psychical Research, 74*(898), 1-14.

Dagnall, N., Drinkwater, K., Denovan, A., Parker, A., & Rowley, K. (2016b). Misperception of chance, conjunction, framing effects and belief in the paranormal: A further evaluation. *Applied Cognitive Psychology, 30*, 409-419.

Dagnall, N., Drinkwater, K., Parker, A., & Rowley, K. (2014). Misperception of chance, conjunction, belief in the paranormal and reality testing: A reappraisal. *Applied Cognitive Psychology, 28*(5), 711-719.

Dagnall, N., Parker, A., & Munley, G. (2007). Paranormal belief and reasoning. *Personality and Individual Differences, 43*(6), 1406-1415.

Dalecki, M.G., Whitehead, J.C., & Blomquist, G.C. (1993). Sample non-response bias and aggregate benefits in contingent valuation: An examination of early, late and non-respondents. *Journal of Environmental Management, 38*(2), 133-143.

Danner, U.N., Sanders, N., Smeets, P.A., van Meer, F., Adan, R.A., Hoek, H.W., & van Elburg, A.A. (2012). Neuropsychological weaknesses in anorexia nervosa: Set-shifting, central coherence, and decision making in currently ill and recovered women. *International Journal of Eating Disorders, 45*(5), 685-694.

Darwin, H., Neave, N., & Holmes, J. (2011). Belief in conspiracy theories. The role of paranormal belief, paranoid ideation and schizotypy. *Personality and Individual Differences, 50*(8), 1289-1293.

Davies, M.F. (1988). Paranormal beliefs in British and Southern USA college students. *Psychological Reports, 62*, 163-166.

Dean, C.E., Akhtar, S., Gale, T.M., Irvine, K., Grohmann, D., & Laws, K.R. (2022). Paranormal beliefs and cognitive function: A systematic review and assessment of study quality across four decades of research. *PloS One, 17*(5), e0267360.

Dean, C.E., Akhtar, S., Gale, T.M., Irvine, K., Wiseman, R., & Laws, K.R. (2021). Development of the Paranormal and Supernatural Beliefs Scale using classical and modern test theory. *BMC Psychology, 9*(1), 1-20.

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical education, 44*(1), 109-117.

De Luccia, G., & Ortiz, K.Z. (2009). Performance of a Brazilian population in the EC 301 calculation and number processing battery: A pilot study. *Arquivos de Neuro-Psiquiatria, 67*, 432-438.

Demirtas, A.S., & Yildiz, B. (2019). Hopelessness and perceived stress: The mediating role of cognitive flexibility and intolerance of uncertainty. *The Journal of Psychiatry and Neurological Sciences, 3*(32), 259-267.

Denovan, A., Dagnall, N., Drinkwater, K., & Parker, A. (2018). Latent profile analysis of schizotypy and paranormal belief: Associations with probabilistic reasoning performance. *Frontiers in Psychology, 9*, 35.

Dennis, J.P., & Vander Wal, J.S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research, 34*(3), 241-253.

de Oliveira Souza, C., Voos, M.C., Francato, D.V., Chien, H.F., & Barbosa, E.R. (2013). Influence of educational status on executive function and functional balance in individuals with Parkinson disease. *Cognitive and Behavioural Neurology, 26*(1), 6-13.

DeVellis, R.F. Classical test theory. *Medical Care*, S50-S59.

Dewan, W.J. (2013). Sceptic culture: Traditions of disbelief in New Mexico. In O. Jenzen, & S.R. Munt (Eds.), *The Ashgate Research Companion to Paranormal Cultures* (pp. 107-120). Routledge.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.

Douw, L., Wakeman, D.G., Tanaka, N., Liu, H., & Stufflebeam, S.M. (2016). State-dependent variability of dynamic functional connectivity between frontoparietal and default networks relates to cognitive flexibility. *Neuroscience, 339*, 12-21.

Downes, M.J., Brennan, M.L., Williams, H.C., & Dean, R.S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open, 6*(12), e011458.

Drinkwater, K., Dagnall, N., Denovan, A., & Parker, A. (2017). An assessment of the dimensionality and factorial structure of the revised paranormal belief scale. *Frontiers in Psychology, 8*, 1693.

Drinkwater, K., Dagnall, N., Denovan, A., & Parker, A. (2018). The Australian Sheep-Goat Scale: An evaluation of factor structure and convergent validity. *Frontiers in Psychology, 9*, 1594.

Drinkwater, K., Dagnall, N., Denovan, A., & Parker, A. (2019). The moderating effect of mental toughness: Perception of risk and belief in the paranormal. *Psychological Reports, 122*(1), 268-287.

Drinkwater, K., Dagnall, N., Grogan, S., & Riley, V. (2017). Understanding the unknown: A thematic analysis of subjective paranormal experiences. *Australian Journal of Parapsychology, 17*(1), 23-46.

Drinkwater, K., Dagnall, N., & Parker, A. (2012). Reality testing, conspiracy theories, and paranormal beliefs. *Journal of Parapsychology, 76*(1), 57-77.

Drinkwater, K., Denovan, A., Dagnall, N., & Parker, A. (2017). An assessment of the dimensionality and factorial structure of the Revised Paranormal Belief Scale. *Frontiers in Psychology, 8*, 1693.

Downing, S. M. (2003). Item response theory: applications of modern test theory in medical education. *Medical education, 37*(8), 739-745.

Dudley, R.E., John, C.H., Young, A.W., & Over, D.E. (1997). Normal and abnormal reasoning in people with delusions. *British Journal of Clinical Psychology, 36*(2), 243-258.

Dudley, R.T. (1999). Effect of restriction of working memory on reported paranormal belief. *Psychological Reports, 84*(1), 313-316.

Duncan, J. (2004). EPS Mid-Career Award 2004: Brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology, 59*(1), 2-7.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179.

Duncan, J., & Owen, A.M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23*(10), 475-483.

Eder, E., Turic, K., Milasowszky, N., Van Adzin, K., & Hergovich, A. (2011). The relationships between paranormal belief, creationism, intelligent design and evolution at secondary schools in Vienna (Austria). *Science & Education, 20*(5), 517-534.

Emmons, C.F., & Sobal, J. (1981). Paranormal beliefs: Testing the marginality hypothesis. *Sociological Focus, 14*(1), 49-56.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Esen-Aygun, H. (2018). The relationship between pre-service teachers' cognitive flexibility and interpersonal problem solving skills. *Eurasian Journal of Educational Research, 77*, 105-128.

Eysenbach, G. (2004). Improving the quality of Web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research, 6*(3), e132.

Eysenck, H.J. (1960). Classification and the problem of diagnosis. In H.J. Eysenck (Ed.), *Handbook of Abnormal Psychology* (pp. 1-31). Pitman.

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

Ferguson, H.J., Brunsdon, V.E., & Bradford, E.E. (2021). The developmental trajectories of executive function from adolescence to old age. *Scientific Reports, 11*(1), 1-7.

Formann, A.K., & Piswanger, K. (1979). *WMT – Wiener Matrizentest.* Beltz Test.

Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry, 36*(6), 717-732.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42.

French, C.C., & Stone, A. (2013). *Anomalistic Psychology: Exploring Paranormal Belief and Experience.* Macmillan Publishers Ltd.

French, C.C., & Wilson, K. (2007). Cognitive factors underlying paranormal beliefs and experiences. In S.D. Sala (Ed.), *Tall tales about the mind and brain: Separating fact from fiction* (pp. 3-22). Oxford University Press.

Freeston, M. H., Rhéaume, J., Letarte, H., Dugas, M. J., & Ladouceur, R. (1994). Why do people worry? *Personality and Individual Differences, 17*(6), 791–802.

Fretwell, C.E., Lewis, C.C., & Hannay, M. (2013). Myers-Briggs Type Indicator, A/B personality types, and locus of control: Where do they intersect? *American Journal of Management, 13*(3), 57-66.

Furr, M. (2011). *Scale Construction and Psychometrics for Social and Personality Psychology*. London: SAGE Publications Ltd.

Gagné, H., & McKelvie, S.J. (1990). Effects of paranormal beliefs on response bias and self-assessment of performance in a signal detection task. *Australian Journal of Psychology, 42*(2), 187-195.

Gallagher, C., Kumar, V.K., & Perala, R.J. (1994). The Anomalous Experiences Inventory: Reliability and validity. *The Journal of Parapsychology, 58*(4), 402-428.

Gannon, M.J., Nothern, J.C., & Carroll, S.J. (1971). Characteristics of nonrespondents among workers. *Journal of Applied Psychology, 55*(6), 586-588.

Genovese, J.E.C. (2005). Paranormal beliefs, schizotypy, and thinking styles among teachers and future teachers. *Personality and Individual Differences, 39*(1), 93-102.

Gianotti, L.R., Mohr, C., Pizzagalli, D., Lehmann, D., & Brugger, P. (2001). Associative processing and paranormal belief. *Psychiatry and Clinical Neurosciences, 55*(6), 595-603.

Gilbert, S.J., & Burgess, P.W. (2008). Executive function. *Current Biology, 18*(3), R110-R114.

Glick, P., Gottesman, D., & Jolton, J. (1989). The fault is not in the stars: Susceptibility of skeptics and believers in astrology to the Barnum effect. *Personality and Social Psychology Bulletin, 15*(4), 572-583.

Gnambs, T. (2017). Human capital and reemployment success: The role of cognitive abilities and personality. *Journal of Intelligence, 5*(1), 9.

Gollan, T.H., Kleinman, D., & Wierenga, C.E. (2014). What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General, 143*(6), 2167-2195.

Göritz, A.S., & Schumacher, J. (2000). The WWW as a research medium: An illustrative survey on paranormal belief. *Perceptual and Motor Skills, 90*(3_suppl), 1195-1206.

Goulding, A. (2004). Schizotypy models in relation to subjective health and paranormal beliefs and experiences. *Personality and Individual Differences, 37*(1), 157-167.

Gow, K., Lane, A., & Chant, D. (2003). Personality characteristics, beliefs, and the near-death experience. *Australian Journal of Clinical and Experimental Hypnosis, 31*(2), 128-152.

Graeupner, D., & Coman, A. (2017). The dark side of meaning-making: How social exclusion leads to superstitious thinking. *Journal of Experimental Social Psychology, 69*, 218-222.

Grant, D.A., & Berg, E. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weight-type card-sorting problem. Journal of Experimental Psychology, 38(4), 404-411.

Gray, S.J., & Gallo, D.A. (2016). Paranormal psychic believers and skeptics: A large-scale test of the cognitive differences hypothesis. *Memory & Cognition, 44*(2), 242-261.

Greenaway, K.H., Louis, W.R., & Hornsey, M.J. (2013). Loss of control increases belief in precognition and belief in precognition increases control. *PloS One, 8*(8), e71327.

Greening, E.K. (2002). *The relationship between false memory and paranormal belief.* [Doctoral dissertation, University of Hertfordshire].

Greenland, S., & O'rourke, K. (2001). On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics, 2*(4), 463-471.

Greenland, S., & Robins, J. (1994). Invited commentary: Ecologic studies – biases, misconceptions, and counterexamples. *American Journal of Epidemiology, 139*(8), 747-760.

Griffiths, O., Shehabi, N., Murphy, R.A., & Le Pelley, M.E. (2019). Superstition predicts perception of illusory control. *British Journal of Psychology, 110*(3), 499-518.

Guerra-Carrillo, B., Katovich, K., & Bunge, S.A. (2017). Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample. *PLoS One, 12*(8), e0182276.

Halsey, L.G. (2019). The reign on the *p*-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters, 15*, 20190174.

Hambleton, R.K., & Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 28-47.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5-25.

Haraldsson, E. (1985). Representative national surveys of psychic phenomena: Iceland, Great Britain, Sweden, USA and Gallup's multinational survey. *Journal of the Society for Psychical Research, 53*(801), 145-158.

Haraldsson, E., & Houtkooper, J.M. (1996). Traditional Christian beliefs, spiritualism, and the paranormal: An Icelandic-American comparison. *The International Journal for the Psychology of Religion, 6*(1), 51-64.

Harari, M.B., Parola, H.R., Hartwell, C.J., & Riegelman, A. (2020). Literature searches in systematic reviews and meta-analyses: A review, evaluation, and recommendations. *Journal of Vocational Behavior, 118*, 103377.

Hart, J.W., Sullivan-Sanchez, C., Packer, T., & Loveless, J. (2013). Is any explanation better than no explanation? Intolerance of uncertainty and paranormal beliefs. *Social Behaviour and Personality: An International Journal, 41*(2), 343-344.

Hartshorne, J.K., & Germine, L.T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26*(4), 433-443.

Hawkins, D.F. (1975). Estimation of nonresponse bias. *Sociological Methods & Research, 3*(4), 461-488.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186.

Henry, L.A. (2001). How does the severity of a learning disability affect working memory performance? *Memory, 9*(4-6), 233-247.

Hergovich, A. (2003). Field dependence, suggestibility, and belief in paranormal phenomena. *Personality and Individual Differences, 34*(2), 195-209.

Hergovich, A., & Arendasy, M. (2005). Critical thinking ability and belief in the paranormal. *Personality and Individual Differences, 38*(8), 1805-1812.

Hergovich, A., & Hörndler, H. (1994). *Gestaltwahrnehmungstest: e. computerbasiertes Verfahren zur Messung d. Feldartikulation; Manual*. Swets Test Services.

Hergovich, A., Schott, R., & Arendasy, M. (2005). Paranormal belief and religiosity. *Journal of Parapsychology, 69*(2), 293-303.

Hergovich, A., Schott, R., & Burger, C. (2010). Biases evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology, 29*(3), 188-209.

Hillstrom, E.L., & Strachan, M. (2000). Strong commitment to traditional Protestant religious beliefs is negatively related to beliefs in paranormal phenomena. *Psychological Reports, 86*(1), 183-189.

Hitchman, G.A., Roe, C.A., & Sherwood, S.J. (2012). A reexamination of nonintentional precognition with openness to experience, creativity, psi beliefs, and luck beliefs as predictors of success. *Journal of Parapsychology, 76*(1), 109-145.

Hoepfl, M.C. (1997). Choosing qualitative research: A primer for technology education researchers. *Journal of Technology Education, 9*(1), 47-63.

Holton III, E.F., Bates, R.A., Bookter, A.I., & Yamkovenko, V.B. (2007). Convergent and divergent validity of the learning transfer system inventory. *Human Resource Development Quarterly, 18*(3), 385-419.

Horton, R. (2002). The hidden research paper. *Jama, 287*(21), 2775-2778.

Hyman, R. (1981). The psychic reading. *Annals of the New York Academy of Sciences, 364*(1), 169-181.

Ilori, O.S., Adebayo, S.O., & Ogunleye, A.J. (2014). Religiosity, paranormal beliefs, and psychopathological symptoms in two ethnic samples. *Humanities and Social Sciences Letters, 2*(4), 192-202.

IMAGE. (2018). Alceste [Computer software]. Retrieved from http://www.image-zafar.com/Logicieluk.html

Imrey, P.B. (2020). Limitations of meta-analyses of studies with high heterogeneity. *JAMA Network Open, 31*(1), e1919325.

Ioannidis, J.P. (2007). Limitations are not properly acknowledged in the scientific literature. *Journal of Clinical Epidemiology, 60*(4), 324-329.

Irwin, H.J. (1993). Belief in the paranormal: A review of the empirical literature. *Journal of the American Society for Psychical Research, 87*(1), 1-39.

Irwin, H.J. (2000). Age and sex differences in paranormal beliefs: A response to Vitulli, Tipton, and Rowe (1999). *Psychological Reports, 86*(2), 595-596.

Irwin, H.J. (2000). Belief in the paranormal and a sense of control over life. *European Journal of Parapsychology, 15*, 68-78.

Irwin, H.J. (2009). *The psychology of paranormal belief: A researcher's handbook.* University of Hertfordshire Press.

Irwin, H.J. (2015). Thinking style and the making of a paranormal disbelief. *Journal of the Society for Psychical Research, 79*(920), 129-139.

Irwin, H.J., Dagnall, N., & Drinkwater, K. (2012). Paranormal belief and biases in reasoning underlying the formation of delusions. *Australian Journal of Parapsychology, 12*(1), 7-21.

Irwin, H.J., Dagnall, N., & Drinkwater, K. (2015). The role of doublethink and other coping processes in paranormal and related beliefs. *Journal of the Society for Psychical Research, 79*(2), 80-96.

Irwin, H.J., Dagnall, N., & Drinkwater, K. (2021). The ethics of belief in paranormal phenomena. *Journal of Anomalous Experience and Cognition (JAEX), 2*(2), 49-79.

Irwin, H.J., Drinkwater, K., & Dagnall, N. (2014). Are believers in the paranormal inclined to jump to conclusions? *Australian Journal of Parapsychology, 14*(1), 69-82.

Irwin, H.J., & Green, M.J. (1999). Schizotypal processes and belief in the paranormal: A multidimensional study. *European Journal of Parapsychology, 14*, 1-5.

Irwin, H.J., & Marks, A.D. (2013). The Survey of Scientifically Unaccepted Beliefs: A new measure of paranormal and related beliefs. *Australian Journal of Parapsychology, 13*(2), 133-167.

Irwin, H.J., Marks, A.D., & Geiser, C. (2018). Belief in the paranormal: A state, or a trait? *Journal of Parapsychology, 82*(1), 24-40.

Jabrayilov, R., Emons, W.H., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied psychological measurement, 40*(8), 559-572.

Jackson, O. (1985). *Manual for the Multidimensional Aptitude Battery*. Research Psychologists Press.

Jackson, J.J., Hill, P.L., Payne, B.R., Parisi, J.M., & Stine-Morrow, E.A.L. (2020). Linking openness to cognitive ability in older adulthood: The role of active diversity. *Aging & Mental Health, 24*(7), 1079-1087.

Johnco, C., Wuthrich, V.M., & Rapee, R.M. (2015). The impact of late-life anxiety and depression on cognitive flexibility and cognitive restructuring skill acquisition. *Depression and Anxiety, 32*, 754-762.

Jones, W., Russell, D., & Nickel, T. (1977). Belief in the paranormal scale: An objective instrument to measure belief in magical phenomena and causes. *JSAS Catalogue of Documents in Psychology, 7*(100), Ms. No. 1577.

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL 8 user's reference guide.* Scientific Software International.

Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Jama, 282*(11), 1054-1060.

Kaiser, R.H., Andrews-Hanna, J.R., Metcalf, C.A., & Dimidjian, S. (2015). Dwell or decenter? Rumination and decentering predict working memory updating after interpersonal criticism. *Cognitive Therapy and Research, 39*(6), 744-753.

Karayagiz, S., & Aktan, T. (2020). Investigation of the relationships between paranormal beliefs, parental bonding, and adult attachment of university students. *Imagination, Cognition and Personality: Consciousness in Theory, Research, and Clinical Practice, 39*(3), 277-290.

Kaufman, A.S., Kaufman, J.C., Liu, X., & Johnson, C.K. (2009). How do educational attainment and gender relate to fluid intelligence, crystallised intelligence, and academic skills at ages 22-90 years? *Archives of Clinical Neuropsychology, 24*(2), 153-163.

Keirsey, D. (1998). *Please understand me II.* Prometheus Nemesis Book Company.

Kennedy, J.E. (2005). Personality and motivations to believe, misbelieve, and disbelieve in paranormal phenomena. *Journal of Parapsychology, 69*, 263-292.

Kennedy, K.M., & Raz, N. (2009). Aging white matter and cognition: Differential effects of regional variations in diffusion properties on memory, executive functions, and speed. *Neuropsychologica, 47*(3), 916-927.

Kim, C., Cilles, S.E., Johnson, N.D., & Gold, B.T. (2011). Domain general and domain preferential brain regions associated with different types of task switching: A meta-analysis. *Human Brain Mapping, 33*(1), 130-142.

Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation.* Sage.

Kontto, J., Tolonen, H., & Salonen, A.H. (2020). What are we missing? The profile of non-respondents in the Finnish gambling 2015 survey. *Scandinavian Journal of Public Health, 48*(1), 80-87.

Kose, I. A., & Demirtasli, N. C. (2012). Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test length and sample size. *Procedia-Social and Behavioral Sciences, 46*, 135-140.

Kossowska, M., Czernatowicz-Kukuczka, A., & Sekerdej, M. (2017). Many faces of dogmatism: Prejudice as a way of protecting certainty against value violators among dogmatic believers and atheists. *British Journal of Psychology, 108*, 127-147.

Krummenacher, P., Mohr, C., Haker, H., & Brugger, P. (2010). Dopamine, paranormal belief, and the detection of meaningful stimuli. *Journal of Cognitive Neuroscience, 22*(8), 1670-1681.

Kruth, J.G. (2015). Five qualitative research approaches and their applications in parapsychology. *The Journal of Parapsychology, 79*(2), 219-233.

Kumar, V.K., Pekala, R.J., & Cummings, J. (1993). Sensation seeking, drug use and reported paranormal beliefs and experiences. *Personality and Individual Differences, 14*(5), 685-691.

Kumar, P., Saini, S.S., Sharma, R., & Kumar, K. (2020). Personality and paranormal belief: A study among university students. *Journal of Psychosocial Research, 15*(2), 139-148.

Kwapil, T.R., & Barrantes-Vidal, N. (2015). Schizotypy: Looking back and moving forward. *Schizophrenia Bulletin, 41*(suppl_2), S366-S373.

Lamont, P., Coelho, C., & Mckinlay, A. (2009). Explaining the unexplained: Warranting disbelief in the paranormal. *Discourse Studies, 11*(5), 543-559.

Lange, R., Irwin, H.J., & Houran, J. (2000). Top-down purification of Tobacyk's revised paranormal belief scale. *Personality and Individual Differences, 29*(1), 131-156.

Lange, R., Irwin, H.J., & Houran, J. (2001). Objective measurement of paranormal belief: A rebuttal to Vitulli. *Psychological Reports, 88*(3), 641-644.

Lange, R., & Thalbourne, M.A. (2002). Rasch scaling paranormal belief and experience: Structure and semantics of Thalbourne's Australian Sheep-Goat Scale. *Psychological Reports, 91*(3_suppl), 1065-1073.

Lannoy, S., Duka, T., Carbia, C., Billieux, J., Fontesse, S., Dormal, V., Gierski, F., López-Caneda, E., Sullivan, E.V., & Maurage, P. (2021). Emotional processes in binge drinking: A systematic review and perspective. *Clinical Psychology Review, 84*, 101971.

Lasikiewicz, N. (2016). Perceived stress, thinking style, and paranormal belief. *Imagination, Cognition and Personality, 35*(3), 306-320.

Lawrence, T.R. (1995). How many factors of paranormal belief are there? A critique of the Paranormal Belief Scale. *The Journal of Parapsychology, 59*(1), 3-26.

Lawrence, T., Edwards, C., Barraclough, N., Church, S., & Hetherington, F. (1995). Modelling childhood causes of paranormal belief and experience: Childhood trauma and childhood fantasy. *Personality and Individual Differences, 19*(2), 209-215.

Lawrence, E., & Peters, E. (2004). Reasoning in believers in the paranormal. *The Journal of Nervous and Mental Disease, 192*(11), 727-733.

Laws, K.R. (2013). Negativland – a home for all findings in psychology. *BMC Psychology, 1*(1), 1-8.

Laws, K.R. (2016). Psychology, replication & beyond. *BMC Psychology, 4*(1), 1-8.

Laws, K.R., Patel, D.D., & Tyson, P.J. (2008). Awareness of everyday executive difficulties precede overt executive dysfunction in schizotypal subjects. *Psychiatry Research, 160*(1), 8-14.

Lelorain, S., Tessier, P., Florin, A., & Bonnaud-Antignac, A. (2012). Posttraumatic growth in long term breast cancer survivors: Relation to coping, social support and cognitive processing. *Journal of Health Psychology, 17*(5), 627-639.

Lesaffre, L., Kuhn, G., Jopp, D.S., Mantzouranis, G., Diouf, C.N., & Rochat, D. (2021). Talking to the dead in the classroom: How a supposedly psychic event impacts beliefs and feelings. *Psychological Reports, 124*(6), 2427-2452.

Lester, D., Thinschmidt, J.S., & Trautman, L.A. (1987). Paranormal belief and Jungian dimensions of personality. *Psychological Reports, 61*(1), 182.

Lieberman, L., Gorka, S.M., Sarapas, C., & Shankman, S.A. (2016). Cognitive flexibility mediates the relation between intolerance of uncertainty and safety signal responding in those with panic disorder. *Cognition and Emotion, 30*(8), 1495-1503.

Linacre, J.M. (2002A). Optimising rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J.M. (2002B). What do infit, outfit, mean-square and standardised mean? *Rasch Measurement Transactions, 16*(2), 878.

Lindeman, M., & Aarnio, K. (2006). Paranormal beliefs: Their dimensionality and correlates. *European Journal of Personality, 20*, 585-602.

Lindeman, M., Riekki, T., & Hood, B.M. (2011). Is weaker inhibition associated with supernatural beliefs? *Journal of Cognition and Culture, 11*(1-2), 231-239.

Lindeman, M., Svedholm-Häkkinen, A.M. (2016). Does poor understanding of the physical world predict religious and paranormal beliefs? *Applied Cognitive Psychology, 30*(5), 736-742.

Lindfors, M., Winberg, M., & Bodin, M. (2019). The role of students' scientific epistemic beliefs in computer-simulated problem solving. *Scandinavian Journal of Educational Research, 63*(1), 124-144.

Lobato, E., Mendoza, J., Sims, V., & Chin, M. (2014). Examining the relationship between conspiracy theories, paranormal beliefs, and pseudoscience acceptance among a university population. *Applied Cognitive Psychology, 28*, 617-625.

Loftus, A.M., Gasson, N., Lopez, N., Sellner, M., Reid, C., Cocks, N., & Lawrence, B.J. (2021). Cognitive reserve, executive function, and memory in Parkinson's disease. *Brain Sciences, 11*, 992.

Lövdén, M., Fratiglioni, L., Glymour, M.M., Lindenberger, U., & Tucker-Drob, E.M. (2020). Education and cognitive functioning across the life span. *Psychological Science in the Public Interest, 21*(1), 6-41.

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment, 1*(1), 1-11.

Mair, P., & Hatzinger, R. (2007). Extended Rasch modelling: The eRm package for the application of IRT models in R. *Journal of Statistical Software, 20*(9), 1-20.

Mair, P., Hatzinger, R., & Maier, M. (2021). *eRm: Extended Rasch modelling*. 1.0-2. Retrieved from https://cran.r-project.org/web/packages/eRm/index.html

Majima, Y. (2015). Belief in pseudoscience, cognitive style and science literacy. *Applied Cognitive Psychology*, 29(4), 552-559.

Mallett, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness, 4*(3), 445-455.

Markovsky, B., & Thye, S.R. (2001). Social influence on paranormal beliefs. *Social Perspectives, 44*(1), 21-44.

Marsiske, M., & Margrett, J.A. (2006). Everyday problem solving and decision making. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging*, (6th ed., pp. 315-342). Academic Press.

Martin, M.M., & Anderson, C.M. (1998). The cognitive flexibility scale: Three validity studies. *Communication Reports*, *11*(1), 1-9.

Martin, M.M., & Rubin, R.B. (1995). A new measure of cognitive flexibility. *Psychological Reports*, *76*, 623-626.

Mason, O.J., & Claridge, G. (2015). *Schizotypy: New dimensions.* Routledge.

Mathijsen, F.P. (2010). Young people and paranormal experiences: Why are they scared? A cognitive pattern. *Archive for the Psychology of Religion*, *32*(3), 345-361.

Mathijsen, F.P. (2012). Adolescents and spiritualism: Is this a good way to cope with fear? A qualitative approach. *Mental Health, Religion & Culture*, *15*(5), 485-494.

Mauzay, D., & Cuttler, C. (2018). Dysfunctional cognitions mediate the relationships between religiosity, paranormal beliefs, and symptoms of obsessive-compulsive disorder. *Mental Health, Religion & Culture*, *21*(8), 838-850.

Maxwell, S.E., Lau, M.Y., & Howard, G.S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487-498.

McLean, C.P., & Miller, N.A. (2010). Changes in critical thinking skills following a course on science and pseudoscience: A quasi-experimental study. *Teaching of Psychology*, *37*(2), 85-90.

Meehl, P.E. (1962). Psychopathology and purpose. In P. Hoch & J. Zubin (Eds.), *The future of psychiatry* (pp. 61-69). Grune and Stratton.

Meiran, N., Diamond, G.M., Toder, D., & Nemets, B. (2011). Cognitive rigidity in unipolar depression and obsessive compulsive disorder: Examination of task switching, Stroop, working memory updating and post-conflict adaptation. *Psychiatry Research, 185*(1-2), 149-156.

Mikušková, E.B. (2018). Conspiracy belief of future teachers. *Current Psychology, 37*(3), 692-701.

Miller, E.K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience, 1*, 59-65.

Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202.

Miyake, A., Emerson, M.J., & Friedman, N.P. (2000B). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language, 21*(2), 169-183.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., & Howerter, A. (2000A). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

Morgan, R.K., & Morgan, D.L. (1998). Critical thinking and belief in the paranormal. *College Student Journal, 32*(1), 135-139.

Morris, N., & Jones, D.M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*(2), 111-121.

Mowen, T.J., Heitkamp, A., & Boman, J. (2022). Paranormal activity: Self-control theory and belief in the paranormal. *Deviant Behaviour*, *43*(6), 728-742.

Munakata, Y., Herd, S.A., Chatham, C.H., Depue, B.E., Banich, M.T., & O'Reilly, R.C. (2011). A unified framework of inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453—459.

Murman, D.L. (2015). The impact of age on cognition. In *Seminars in Hearing* (Vol. 36, No. 03, pp. 111-121). Thieme Medical Publishers.

Musch, J., & Ehrenberg, K. (2002). Probability misjudgement, cognitive ability, and belief in the paranormal. *British Journal of Psychology*,*93*(2), 169-177.

Myers, I.B. (1962). *The Myers-Briggs Type Indicator: Manual (1962).* Consulting Psychologists Press.

Newby, R.W., & Davis, J.B. (2004). Relationships between locus of control and paranormal beliefs. *Psychological Reports*, *94*, 1261-1266.

Norris, P., & Epstein, S. (2011). An experimental thinking style: Its facets and relations with objective and subjective criterion measures. *Journal of Personality*, *79*(5), 1043-1080.

Nyberg, L., Magnussen, F., Lundquist, A., Baaré, W., Bartréz-Faz, D., Bertram, L., Boraxbekk, C.J., Brandmaier, A.M., Drevon, C.A., Ebmeier, K., Ghisletta, P., Henson, R.N., Junqué, C., Kievit, R., Kleemeyer, M., Knights, E., Kühn, S., Lindenberger, U., Penninx, B.W.J.H., ... & Fjell, A.M. (2021). Educational attainment does not influence brain aging. *Proceedings of the National Academy of Sciences*, *118*(18), e2101644118.

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petersson, K.M., & Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologica*, *41*(3), 371-377.

Ofori, P.K., Tod, D., & Lavallee, D. (2017). Predictors of superstitious beliefs. *Journal of Psychology in Africa, 27*(1), 1-12.

Orenstein, A. (2002). Religion and paranormal belief. *Journal for the Scientific Stud of Religion, 41*(2), 301-311.

Osborne, J.W., & Fitzpatrick, D.C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research, and Evaluation, 17*(1), 15.

Otis, L.P., & Kuo, E.C.Y. (1984). Extraordinary beliefs among students in Singapore and Canada. *The Journal of Psychology, 116*, 215-226.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*(6), 972-987.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., … & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), 1-11.

Palmer, J., Mohr, C., Krummenacher, P., & Brugger, P. (2007). Implicit learning of sequential bias in a guessing task: Failure to demonstrate effects of dopamine administration and paranormal belief. *Consciousness and Cognition, 16*(2), 498-506.

Panikratova, Y.R., Vlasova, R.M., Akhutina, T.V., Korneev, A.A., Sinitsyn, V.E., & Pechenkova, E.V. (2020). Functional connectivity of the dorsolateral prefrontal cortex contributes to different components of executive functions. *International Journal of Psychophysiology, 151*, 70-79.

Partridge, C. (2013). Haunted Culture: The persistence of belief in the paranormal. In O. Jenzen & S.R. Munt (Eds.), *The Ashgate Research Companion to Paranormal Cultures* (pp. 39-49). Routledge.

Pechey, R., & Halligan, P. (2011). The prevalence of delusion-like beliefs relative to sociocultural beliefs in the general population. *Psychopathology, 44*(2), 106-115.

Peirce, C.S. (1931). Volume 5. In C. Hartshorne & P. Weiss (Eds.), *Collected Papers of Charles Sanders Peirce* (Vols. 1-6). Harvard University Press. (Original work published 1884-1886).

Peltzer, K. (2003). Magical thinking and paranormal beliefs among secondary and university students in South Africa. *Personality and Individual Differences, 35*(6), 1419-1426.

Pérez Navarro, J.M., & Martínez Guerra, X. (2020). Personality, cognition, and morbidity in the understanding of paranormal belief. *PsyCh Journal, 9*(1), 118-131.

Perkins, S.L., & Allen, R. (2006). Childhood physical abuse and differential development of paranormal belief systems. *The Journal of Nervous and Mental Disease, 194*(5), 349-355.

Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: The 21-item Peters et al. Delusions Inventory (PDI). *Schizophrenia Bulletin, 30*(4), 1005-1022.

Pillemer, D., Wink, P., DiDonato, T., & Sanborn, R. (2003). Gender differences in autobiographical memory styles of older adults. *Memory, 11*(6), 525-532.

Pizzagalli, D., Lehmann, D., & Brugger, P. (2001). Lateralized direct and indirect semantic priming effects in subjects with paranormal experiences and beliefs. *Psychopathology, 34*(2), 75-80.

Poreh, A.M., Avital, R., Dines, P.L., & Levin, J.B. (2015). The effects of age of language acquisition on verbal memory tests in a sample of older adults immigrants. *Psychology & Neuroscience, 8*(1), 66-74.

Price, M. (2012). Epidemiology. In P. Wright, J. Stern, & M. Phelan (Eds.), *Core Psychiatry* (3rd Ed.). Elsevier.

Prike, T., Arnold, M.M., & Williamson, P. (2017). Psychics, aliens, or experience? Using the Anomalistic Belief Scale to examine the relationship between type of belief and probabilistic reasoning. *Consciousness and Cognition, 53*, 151-164.

Prike, T., Arnold, M.M., & Williamson, P. (2018). The relationship between anomalistic belief and biases of evidence integration and jumping to conclusions. *Acta Psychologica, 190*, 217-227.

Prince, M. (2012). Epidemiology. In P. Wright, J. Stern, & M. Phelan (Eds.), *Core Psychiatry* (pp.115-129). Elsevier.

Rabeyron, T., & Watt, C. (2010). Paranormal experiences, mental health and mental boundaries, and psi. *Personality and Individual Differences, 48*(4), 487-492.

Radin, D., Nelson, R., Dobyns, Y., & Houtkooper, J. (2006). Reexamining psychokinesis: Comment on Bösch, Steinkamp, and Boller (2006). *Psychological Bulletin, 132*(4), 529-532.

Rado, S. (1953). Dynamics and classification of disordered behaviour. *American Journal of Psychiatry, 110*(6), 406-416.

Raine, A., & Benishay, D. (1995). The SPQ-B: A brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders, 9*(4), 346-355.

Rajesh, A., Noice, T., Noice, H., Jahn, A., Daugherty, A.M., Heller, W., & Kramer, A.F. (2021). Can a theatre acting intervention enhance inhibitory control in older adults? A brain-behaviour investigation. *Frontiers in Human Neuroscience, 15*, 583220.

Raven, J.S. (1976). *The Advanced Progressive Matrices Set 1.* Oxford Psychologists Press.

Raven, J., Raven, J.C., & Court, J.H. (2000). *Manual for standard progressive matrices 2000 edition.* Oxford Psychologists Press.

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Research BMG. (2017). *BMG Halloween Poll: A third of Brits believe in ghosts, spirits, or other types of paranormal activity* [Internet]. BMG Research. URL https://www.bmgresearch.co.uk /bmg-halloween-poll-third-brits-believe-ghosts-spirits-types-paranormal-activity/

Ridolfo, H., Baxter, A., & Lucas, J.W. (2010). Social influences on paranormal belief: Popular versus scientific support. *Current Research in Social Psychology, 15*(3), 33-41.

Riekki, T., Lindeman, M., Aleneff, M., Halme, A., & Nuortimo, A. (2013). Paranormal and religious believers are more prone to illusory face perception than skeptics and non-believers. *Applied Cognitive Psychology, 27*(2), 150-155.

Rizeq, J., Flora, D.B., & Toplak, M.E. (2020). An examination of the underlying dimensional structure of three domains of contaminated mindware: Paranormal beliefs, conspiracy beliefs, and anti-science attitudes. *Thinking & Reasoning*.

Roberts, M.E., Barthel, F.M.S., Lopez, C., Tchanturia, K., & Treasure, J.L. (2011). Development and validation of the Detail and Flexibility Questionnaire (DFlex) in eating disorders. *Eating Behaviours, 12*, 168-174.

Roberts, M.J., & Seager, P.B. (1999). Predicting belief in paranormal phenomena: A comparison of conditional and probabilistic reasoning. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 13*(5), 443-450.

Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., Manes, F., & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain, 133*(1), 234-247.

Roe, C. (1998). Belief in the paranormal and attendance at psychic readings. *Journal of the American Society for Psychical Research, 92*(1), 25-51.

Roe, C.A. (1999). Critical thinking and belief in the paranormal: A re-evaluation. *British Journal of Psychology, 90*(1), 85-98.

Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814.

Rogers, P., Davis, T., & Fisk, J. (2009). Paranormal belief and susceptibility to the conjunction fallacy. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 23*(4), 524-542.

Rogers, P., Fisk, J.E., & Lowrie, E. (2016). Paranormal believers' susceptibility to confirmatory vs disconfirmatory conjunctions. *Applied Cognitive Psychology*, *30*(4), 628-634.

Rogers, P., Fisk, J.E., & Lowrie, E. (2018). Paranormal belief, thinking style preference and susceptibility to confirmatory conjunction errors. *Consciousness and Cognition*, *65*, 182-196.

Rogers, P., Hattersley, M., & French, C.C. (2019). Gender role orientation, thinking style preference and facets of adult paranormality: A mediation analysis. *Consciousness and Cognition*, *76*, 102821.

Rogers, P., & Lowrie, E.L. (2018). Predicting facets of adult paranormality from childhood exposure to inappropriate styles of parenting. *Imagination, Cognition and Personality: Consciousness in Theory, Research, and Clinical Practice*, *38*(2), 139-172.

Rogers, P., Qualter, P., & Phelps, G. (2007). The mediating and moderating effects of loneliness and attachment style on belief in the paranormal. *European Journal of Parapsychology*, *22*(2), 138-165.

Rogers, P., Qualter, P., Phelps, G., & Gardner, K. (2006). Belief in the paranormal, coping and emotional intelligence. *Personality and Individual Differences*, *41*(6), 1089-1105.

Roof, W.C. (1993). *A generation of seekers: The spiritual journeys of the baby boom generation.* Harper.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641.

Ross, M., & Holmberg, D. (1990). Recounting the past: Gender differences in recall of events in the history of a close relationship. In J.M. Olson & M. P. Zanna (Eds.), *Self-inference Processes: The Ontario Symposium, Volume 6*. Lawrence Erlbaum Associates.

Rosseel, Y. (2012). lavaan: An R package for structural equation modelling. *Journal of Statistical Software, 48*(2), 1-36.

Royalty, J. (1995). The generalizability of critical thinking: Paranormal beliefs versus statistical reasoning. *The Journal of Genetic Psychology, 156*(4), 477-488.

Rudski, J. (2004). The illusion of control, superstitious belief, and optimism. *Current Psychology, 22*(4), 306-315.

Rusch, T., Lowry, P.B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management, 54*(2), 189-203.

Sá, W.C., West, R.F., & Stanovich, K.E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*(3), 497-510.

Saher, M., & Lindeman, M. (2005). Alternative medicine: A psychological perspective. *Personality and Individual Differences, 39*(6), 1169-1178.

Salthouse, T.A., Atkinson, T.M., & Berish, D.E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*(4), 566.

Schienle, A., Vaitl, D., & Stark, R. (1996). Covariation bias and paranormal belief. *Psychological Reports, 78*(1), 291-305.

Scott, W.A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry, 25*(4), 405-414.

Sharkness, J., & DeAngelo, L. (2011). Measuring student involvement: A comparison of classical test theory and item response theory in the construction of scales from student surveys. *Research in Higher Education, 52*(2), 480-507.

Silverman, M.J. (2007). The effect of paired pitch, rhythm, and speech on working memory as measured by sequential digit recall. *Journal of Music Therapy, 44*(4), 415-427.

Simmonds-Moore, C. (2014). Exploring the perceptual biases associated with believing and disbelieving in paranormal phenomena. *Consciousness and Cognition, 28,* 30-46.

Simms, L.J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass, 2*(1), 414-433.

Smith, M.D., Foster, C.L., & Stovin, G. (1998). Intelligence and paranormal belief: Examining the role of context. *The Journal of Parapsychology, 62*(1), 65-78.

Smith, C.L., Johnson, J.L., & Hathaway, W. (2009). Personality contributions to belief in paranormal phenomena. *Individual Differences Research, 7*(2), 85-96.

Song, F., Parekh-Bhurke, S., Hooper, L., Loke, Y.K., Ryder, J.J., Sutton, A.J., Hing, C.B., & Harvey, I. (2009). Extent of publication bias in different categories of research cohorts: A meta-analysis of empirical studies. *BMC Medical Research Methodology, 9*(1), 1-14.

Sparks, G.G., & Miller, W. (2001). Investigating the relationship between exposure to television programs that depict paranormal phenomena and beliefs in the paranormal. *Communication Monographs, 68*(1), 98-113.

Sparks, G.G., Nelson, C.L., & Campbell, R.G. (1997). The relationship between exposure to televised messages about paranormal phenomena and paranormal beliefs. *Journal of Broadcasting & Electronic Media, 41*(3), 345-359.

Sparks, G.G., Pellechia, M., & Irvine, C. (1998). Does television news about UFOs affect viewer's UFO beliefs? An experimental investigation. *Communication Quarterly, 46*(3), 284-294.

Spiro, R.J., Coulson, R.L., Feltovich, P.J., & Anderson, D.K. (1988). *Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains* (Technical Report No. 441). Center for the Study of Reading.

Ståhl, T., & Van Prooijen, J.W. (2018). Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. *Personality and Individual Differences, 122*, 155-163.

Stanovich, K.E., & West, R.F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*(2), 342-357.

Stark, R., & Bainbridge, W.S. (1980). Networks of faith: Interpersonal bonds and recruitment to cults and sects. *American Journal of Sociology, 85*(6), 1376-1395.

Steffens, M., Meyhöfer, I., Fassbender, K., Ettinger, U., & Kambeitz, J. (2018). Association of schizotypy with dimensions of cognitive control: A meta-analysis. *Schizophrenia Bulletin, 44*(suppl_2), S512-S524.

Stevens, J. (2012). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

Stone, A. (2016). Rational thinking and belief in psychic abilities: It depends on the level of involvement. *Psychological Reports, 118*(1), 74-89.

Stroebe, W., Gadenne, V., & Nijstad, B.A. (2018). Do our psychological laws apply only to college students? External validity revisited. *Basic and Applied Social Psychology, 40*(6), 384-395.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Stuart-Hamilton, I., Nayak, L., & Priest, L. (2006). Intelligence, belief in the paranormal, knowledge of probability and aging. *Educational Gerontology, 32*(3), 173-184.

Svedholm, A.M., & Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology, 104*(3), 303-319.

Swami, V., Chamorro-Premuzic, T., & Shafi, M. (2009). Psychology in outerspace: Personality, individual difference, and demographic predictors of beliefs about extraterrestrial life. *European Psychologist, 15*(3), 220-228.

Tang, Y., Horikoshi, M., & Li, W. (2016). ggfortify: Unified interface to visualise statistical results of popular R packages. *R Journal, 8*(2), 478-489.

Terhune, D.B., & Smith, M.D. (2006). The induction of anomalous experiences in a mirror-gazing facility: Suggestion, cognitive perceptual personality traits and phenomenological state effects. *The Journal of Nervous and Mental Disease, 194*(6), 415-421.

Tiego, J., Testa, R., Bellgrove, M.A., Pantelis, C., & Whittle, S. (2018). A hierarchical model of inhibitory control. *Frontiers in Psychology, 9*, 1339.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5).

Thalbourne, M.A. (1995). Further studies of the measurement and correlates of belief in the paranormal. *Journal of the American Society for Psychical Research, 89*(3), 233-247.

Thalbourne, M.A. (2010). The Australian Sheep-Goat Scale: Development and empirical findings. *Australian Journal of Parapsychology, 10*(1), 5-39.

Thalbourne, M.A., & Delin, P.S. (1993). A new instrument for measuring the sheep-goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research, 59*, 172-186.

Thalbourne, M.A., Dunbar, K.A., & Delin, P.S. (1995). An investigation into correlates of belief in the paranormal. *Journal of the American Society for Psychical Research, 89*(3), 215-231.

Tisserand, D.J., & Jolles, J. (2003). On the involvement of prefrontal networks in cognitive ageing. *Cortex, 39*(4-5), 1107-1128.

Tobacyk, J. (1983). Cognitive complexity and paranormal beliefs. *Psychological Reports, 52*(1), 101-102.

Tobacyk, J. (1984). Paranormal belief and college grade point average. *Psychological Reports, 54*(1), 217-218.

Tobacyk, J.J. (2004). A revised paranormal belief scale. *The International Journal of Transpersonal Studies, 23*(23), 94-98.

Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena: Assessment instrument development and implications for personality functioning. *Journal of Personality and Social Psychology, 44*(5), 1029-1037.

Tobacyk, J., Milford, G., Springer, T., & Tobacyk, Z. (1988). Paranormal beliefs and the Barnum effect. *Journal of Personality Assessment, 52*(4), 737-739.

Tolonen, H., Dobson, A., Kulathinal, S., & WHO Monica Project. (2005). Effect on trend estimates of the difference between survey respondents and non-respondents: Results from 27 populations in the WHO MONICA Project. *European Journal of Epidemiology, 20*(11), 887-898.

Tosyali, F., & Aktas, B. (2021). Does training analytical thinking decrease superstitious beliefs? Relationship between analytical thinking, intrinsic religiosity, and superstitious beliefs. *Personality and Individual Differences, 183*, 111122.

Toutountzidis, D., Gale, T.M., Irvine, K., Sharma, S., & Laws, K.R. (2022). Childhood trauma and schizotypy in non-clinical samples: A systematic review and meta-analysis. *Plos One, 17*(6), e0270494.

Urbina, S. (2014). *Essentials of psychological testing*. John Wiley & Sons.

Valadez, M.D.L.D., López-Aymes, G., Ruvalcaba, N.A., Flores, F., Ortíz, G., Rodríguez, C., & Borges, Á. (2020). Emotions and reactions to the confinement by COVID-19 of children and adolescents with high abilities and community samples: A mixed methods research study. *Frontiers in Psychology*, 2881.

Vallee, J.M. (2008). Effects of social influence and persuasion on paranormal beliefs. *Journal of Undergraduate Psychological Research, 3*, 47-51.

Van den Bulk, J., & Custers, K. (2010). Belief in complementary and alternative medicine related to age and paranormal beliefs in adults. *European Journal of Public Health, 20*(2), 227-230.

van der Wal, R.C., Sutton, R.M., Lange, J., & Braga, J.P. (2018). Suspicious binds: Conspiracy thinking and tenuous perceptions of causal connections between co-occurring and spuriously correlated events. *European Journal of Social Psychology, 48*(7), 970-989.

Van Elk, M. (2013). Paranormal believers are more prone to illusory agency detection than sceptics. *Consciousness and Cognition, 22*(3), 1041-1046.

Van Elk, M. (2015). Perceptual biases in relation to paranormal and conspiracy beliefs. *PLoS One, 10*(6), e0130422.

Van Elk, M. (2017). The self-attribution bias and paranormal beliefs. *Consciousness and Cognition, 49*, 313-321.

van Prooijen, J.W., Douglas, K.M., & De Inocencio, C. (2018). Connecting the dots: Illusory pattern perception preducts belief in conspiracies and the supernatural. *European Journal of Social Psychology, 48*(3), 320-335.

Van Strien, J.W., Weber, R.F.A., Burdorf, A., & Bangma, C. (2009). Higher free testosterone level is associated with faster visual processing and more flanker interference in older men. *Psychoneuroendocrinology, 34*(4), 546-554.

Vitulli, W.F. (2000). Rejoinder to Irwin's (2000) "Age and sex differences in paranormal beliefs: A response to Vitulli, Tipton, and Rowe (1999)". *Psychological Reports, 87*(2), 699-700.

Vitulli, W.F., Tipton, S.M., & Rowe, J.L. (1999). Beliefs in the paranormal: Age and sex differences among elderly persons and undergraduate students. *Psychological Reports, 85*(3), 847-855.

von Hippel, W., & Gonsalkorale, K. (2005). "That is bloody revolting!" Inhibitory control of thoughts better left unsaid. *Psychological Science, 16*(7), 497-500.

Voos, M.C., Custódio, E.B., & Malaquias Jr, J. (2011). Relationship of executive function and educational status with functional balance in older adults. *Journal of Geriatric Physical Therapy, 34*(1), 11-18.

Voracek, M. (2009). Who wants to believe? Associations between digit ratio (2D:4D) and paranormal and superstitious beliefs. *Personality and Individual Differences, 47*(2), 105-109.

Wagenmakers, E.J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q.F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., ... & Morey, R.D. (2018a). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25(1)*, 58-76.

Wagenmakers, E.J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, L., Selker, R., Gronau, Q.F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J.N., & Morey, R.D. (2018b). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35-57.

Wain, O., & Spinella, M. (2007). Executive functions in morality, religion, and paranormal beliefs. *International Journal of Neuroscience, 117*(1), 135-146.

Wang, C., Zhang, Z., Wiley, J.A., Fu, T., & Yan, J. (2022). Gender differences in pleasure: The mediating roles of cognitive flexibility and emotional expressivity. *BMC Psychiatry*, *22*(1), 1-8.

Ward, S.J., & King, L.A. (2020). Examining the roles of intuition and gender in magical beliefs. *Journal of Research in Personality*, *86*, 103956.

Watson, G. (1980). *Watson-Glaser critical thinking appraisal.* Psychological Corporation.

Watson, G., & Glaser, E.M. (1964). *Critical Thinking Appraisal*. Brace & World.

Watt, C., Watson, S., & Wilson, L. (2007). Cognitive and psychological mediators of anxiety: Evidence from a study of paranormal belief and perceived childhood control. *Personality and Individual Differences*, *42*(2), 335-343.

Wechsler, D. (1955). Wechsler adult intelligence scale. *Archives of Clinical Neuropsychology.*

Werner, S., Praxedes, M., & Kim, H.G. (2007). The reporting of nonresponse analyses in survey research. *Organisational Research Methods*, *10*(2), 287-295.

Wierzbicki, M. (1985). Reasoning errors and belief in the paranormal. *The Journal of Social Psychology*, *125*(4), 489-494.

Wiggins, B.J., & Christopherson, C.D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202-217.

Willard, A.K., & Norenzayan, A. (2013). Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose. *Cognition*, *129*(2), 379-391.

William, D. (2022). How should educational research respond to the replication "crisis" in the social sciences? Reflections on the papers in the Special Issue. *Educational Research and Evaluation, 27*(1-2), 208-214.

Williams, C., Denovan, A., Drinkwater, K., & Dagnall, N. (2022). Thinking style and paranormal belief: The role of cognitive biases. *Imagination, Cognition and Personality: Consciousness in Theory, Research, and Clinical Practice, 41*(3), 274-298.

Williams, L.M., & Irwin, H.J. (1991). A study of paranormal belief, magical ideation as an index of schizotypy and cognitive style. *Personality and Individual Differences, 12*(12), 1339-1348.

Williams, E., & Roberts, B.L.H. (2016). The relationship between paranormal belief and the HEXACO domains of personality. *Journal of Empirical Theology, 29*(2), 212-238.

Wilson, J.A. (2018). Reducing pseudoscientific and paranormal beliefs in university students though a course in science and critical thinking. *Science & Education, 27*(1), 183-210.

Wilson, M.S., Bulbulia, J., & Sibley, C.G. (2014). Differences and similarities in religious and paranormal beliefs: A typology of distinct faith signatures. *Religion, Brain & Behaviour, 4*(2), 104-126.

Wilson, K., & French, C.C. (2006). The relationship between susceptibility to false memories, dissociativity, and paranormal belief and experience. *Personality and Individual Differences, 41*(8), 1493-1502.

Wilson, C.R.E., Gaffan, D., Browning, P.G.F., & Baxter, M.G. (2010). Functional localisation within the prefrontal cortex: Missing the forest for the trees? *Trends in Neurosciences, 33*(12), 533-540.

Wuthrich, V., & Bates, T.C. (2005). Reliability and validity of two Likert versions of the Schizotypal Personality Questionnaire (SPQ). *Personality and Individual Differences, 38*(7), 1543-1548.

Žeželj, I., Pavlović, M., Vladisavljević, M., & Radivojević, B. (2009). Construction and behavioural validation of superstition scale. *Psihologija, 42*(2), 141-158.

Zimmer, T.A. (1985). Belief in UFOs as alternative reality, cultural rejection or disturbed psyche. *Deviant Behaviour, 6*(4), 405-419.

Zmigrod, L., Rentfrow, P.J., Zmigrod, S., & Robbins, T.W. (2019). Cognitive flexibility and religious disbelief. *Psychological Research, 83*, 1749-1759.

Appendices

Appendix A – Table of papers excluded from the systematic literature review (participants < 18)

Study	Sample Size (% women)	Age Range, *M* (SD)	Area	Tests Used	Key Findings
Rogers et al. (2018)	261 (54.0)	16-84, 37.10 (16.50)	Ts	REI, SJQ	- corr. analytical thinking and: ESP ($r = -.24$, $p < .001$), PK ($r = -.29$, $p < .001$), and LAD ($r = -.24$, $p < .001$) subscales of ASGS + corr. intuitive thinking and: ESP ($r = .30$, $p < .001$), PK ($r = .24$, $p < .001$), and LAD ($r = .34$, $p < .001$) subscales of ASGS + corr. total number of conjunction errors and: ESP ($r = .24$, $p < .001$), PK ($r = .19$, $p < .001$), and LAD ($r = .22$, $p < .001$) subscales of ASGS Generation of conjunction errors predicted by: ESP (*exp(b)* = 1.32, $p < .001$), PK (*exp(b)* = 3.16, $p = .003$), and LAD (*exp(b)* = 1.27, $p < .001$) subscales of ASGS
Dagnall et al. (2017)	222 (72.0)	16-63, 30.77 (11.74)	CPb	IPO-RT (RT subscale)	+ corr. paranormal belief and RT scores for: total paranormal belief ($r = .41$, $p < .001$), NAP ($r = .37$, $p < .001$), and TPB ($r = .38$, $p < .001$) subscales of the RPBS - believers demonstrated lower RT ability

					RT deficits had positive effect on TPB (β = .54, $p < .001$) and NAP (β = .48, $p < .001$)
Rogers et al. (2017)	130 (57.3)	16-72, 34.40 (13.30)	CPb	SJQ	+ corr. PK subscale of ASGS and: total conjunction errors ($r = .22$, $p < .05$), and errors for related constituent pairings ($r = .19$, $p < .05$) PK predicted the making of conjunction errors ($exp(b) = 1.26$, $p = .008$)
Lindeman et al. (2015)	2789 (65.0)	15-69, 28.00 (8.87)	CPb	CKCS, teleology statements	+ corr. paranormal beliefs and: ontological confusions ($r = .46$, $p < .05$), and teleology ($r = .31$, $p < .05$) Ontological confusions ($\beta = .41$) and teleology ($\beta = .15$) significantly predicted paranormal beliefs
Pennycook et al. (2012)	53 (72.5)	16-69, 35.04 (12.77)	Ts	CRT, BRC problems, WordSum, BRN problems	- corr. paranormal beliefs and: BRN ($r = -.13$, $p < .05$), CRT ($r = -.31$, $p < .05$), and BRC ($r = -.23$, $p < .05$) <u>Ns.</u> corr. paranormal beliefs and WordSum Paranormal beliefs negatively predicted analytical cognitive style ($\beta = -.20$, $p < .002$)
Rogers et al. (2011)	167 (49.7)	16-71, 32.00 (13.40)	CPb	SJQ	Significant main effect of belief on number of conjunction errors ($F(1,135) = 6.40$, $p = .013$, *partial* $\eta^2 = .05$) – believers made more conjunction errors than sceptics
Lindeman & Aarnio (2007)	239 (77.0)	16-47, 24.20 (/)	Ts	REI, ontological confusion statements	+ corr. intuitive thinking and belief in: paranormal agents ($r = .49$, $p < .001$), paranormal abilities ($r = .51$, $p < .001$), luck ($r = .43$, $p < .001$), astrology ($r = .54$, $p < .001$), and feng shui ($r = .51$, $p < .001$)

					- corr. analytical thinking and belief in: paranormal agents (r = -.21, $p < .001$), paranormal abilities (r = -.19, $p < .01$), luck (r = -.22, $p < .001$), astrology (r = -.28, $p < .001$), and feng shui (r = -.27, $p < .001$) + corr. all six types of ontological confusions and all five paranormal beliefs (all rs ≥ .30, ps < .001)
Rogers et al. (2006)	253 (38.7)	17-82, 38.50 (12.50)	O	SREIT, WCQ	- corr. paranormal beliefs and all four SREIT subscales: mood regulation (r = -.17, $p < .01$), appraisal of emotions (r = -.17, $p < .01$), social skills (r = -.15, $p < .05$), utilisation of emotions (r = -.16, $p < .01$) Ns. corr. active-cognitive coping and paranormal beliefs Ns. neither active-cognitive coping or SREIT scores predicted paranormal beliefs
Lindeman & Aarnio (2006)	3261 (74.0)	15-60, 24.00 (4.67)	Ts	REI	Intuitive thinking positively predicted higher-order paranormal beliefs (β = 0.25, $p < .01$), while analytical thinking negatively predicted higher-order paranormal beliefs (β = -0.17, $p < .001$)
Aarnio & Lindeman (2005)	3141 (74.0)	16-60, 24.00 (4.50)	Ts	REI	+ corr. paranormal beliefs and intuitive thinking (r = .34, $p < .001$) - corr. paranormal beliefs and analytical thinking (r = -.14, $p < .001$)
Saher & Lindeman (2005)	3261 (74.0)	15-60, 24.00 (/)	Ts	REI	+ corr. paranormal beliefs and intuitive thinking (r = .34, $p < .001$) - corr. paranormal beliefs and analytical thinking (r = -.15, $p < .001$)
Farias et al. (2005)	99 (56.6)	17-79, 38.20 (21.10)	CPb	Visual perception task	+ corr. paranormal beliefs and complex dot patterns for: total complex reports (r = .29, $p < .01$), number of different complex reports (r = .25, $p < .01$) - corr. paranormal beliefs and complex dot patterns for: latency of first complex report (r - .24, $p < .05$)

					Ns. corr. paranormal beliefs and simple dot patterns for: total simple reports, number of different simple reports, or latency of first simple report
Roig et al. (1998)	814 (54.8)	17-47, 20.40 (/)	Ts	IBI	Paranormal believers scored significantly higher compared to sceptics for: global irrational thinking ($F(1,407) = 18.24$, $p < .001$), rigidity ($F(1,407) = 15.38$, $p < .001$), and worrying ($F(1,407) = 18.24$, $p < .001$)

Note: Ts = Thinking Style, CPb = Cognitive and Perceptual Biases, O = Other Cognitive Functions, REI = Rational and Experiential Inventory (Epstein et al., 1996), SJQ = Scenario Judgements Questionnaire (Rogers et al., 2016; Rogers et al., 2011), IPO-RT = Inventory of Personality Organization (Lenzenweger et al., 2001), RT = reality testing, ASGS = Australian Sheep-Goat Scale (Thalbourne & Delin, 1993), ESP = extrasensory perception, LAD = life after death, PK = psychokinesis, NAP = new age philosophy, TPB = traditional paranormal beliefs, RPBS = Revised Paranormal Belief Scale (Tobacyk, 2004; Lange et al., 2000), CKCS = Core Knowledge Confusions scale (Lindeman & Aarnio, 2007; Lindeman et al., 2008), CRT = Cognitive Reflection Test (Frederick, 2005), BRC = base-rate conflict, BRN = base-rate neutral, SREIT = Self-Report Emotional Intelligence Test (Schutte et al., 1998), WCQ = Ways of Coping Questionnaire (Folkman & Lazarus, 1988), IBI = Irrational Beliefs Inventory (Koopmans et al., 1994)

Appendix B – Table showing alternate categorisations of studies included in the systematic review

	Section					
Study	Perceptual & cognitive biases	Reasoning	Intelligence, critical thinking & academic performance	Thinking style	Executive function & memory	Other cognitive functions
Alcock & Otis (1980)	✓		✓̲			
Andrews & Tyson (2018)			✓̲			
Barberia et al. (2018)	✓̲					
Betsch et al. (2020)	✓		✓̲	✓		
Blackmore (1997)		✓̲			✓	
Blackmore & Moore (1994)	✓̲				✓	
Blanco et al. (2015)	✓̲					
Branković (2019)	✓			✓̲		
Bressan (2002)		✓̲			✓	
Brugger et al. (1990)		✓̲			✓	

Brugger et al. (1991)	✓	✓̲			✓
Caputo (2017)	✓̲				
Dagnall et al. (2007)		✓̲			✓
Dagnall et al. (2014)		✓̲			✓
Dagnall et al. (2016A)		✓̲			✓
Dagnall et al. (2016B)		✓̲			✓
Denovan et al. (2018)		✓̲			✓
Drinkwater et al. (2019)	✓̲				
Dudley (1999)					✓̲
Gagné & McKelvie (1990)	✓̲	✓			✓
Genovese (2005)				✓̲	
Gianotti et al. (2001)	✓			✓̲	
Gray & Gallo (2016)			✓	✓	✓̲
Greening (2002)					✓̲
Griffiths et al. (2019)	✓̲	✓			✓

Hergovich (2003)	✓			✓̲	✓	
Hergovich & Arendasy (2005)		✓	✓̲		✓	
Irwin (2015)				✓̲		
Irwin & Green (1998-99)					✓	✓̲
Irwin et al. (2014)	✓̲	✓			✓	
Krummenacher et al (2010)	✓̲				✓	
Lasikiewicz (2016)				✓̲		
Lawrence & Peters (2004)		✓̲			✓	
Lesaffre et al. (2020)	✓̲	✓				
Lindeman & Svedholm-Häkkinen (2016)				✓̲	✓	
Lindeman et al. (2011)					✓̲	
Majima (2015)			✓	✓̲		
McLean & Miller (2010)			✓̲			

Mikušková & Cavojavá (2020)	✓	✓		✓̲		
Morgan & Morgan (1998)			✓̲			
Musch & Ehrenberg (2002)		✓̲	✓		✓	
Palmer et al. (2007)		✓			✓	✓̲
Pérez-Navarro & Martínez-Guerra (2020)		✓̲			✓	
Pizzagalli et al. (2001)						✓̲
Prike et al. (2017)		✓̲			✓	
Prike et al. (2018)	✓̲	✓			✓	
Riekki et al. (2013)	✓̲				✓	
Rizeq et al. (2020)			✓	✓̲	✓	
Roberts & Seager (1999)		✓̲			✓	
Roe (1999)			✓̲			
Rogers et al. (2009)		✓̲			✓	

Rogers et al. (2016)	✓	✓̲			✓	
Rogers et al. (2019)				✓̲		
Royalty (1995)		✓	✓̲		✓	
Rudski (2004)	✓̲	✓			✓	
Schienle et al. (1996)	✓̲	✓			✓	
Simmonds-Moore (2014)	✓̲				✓	
Smith et al. (1998)			✓̲		✓	
Ståhl & van Prooijen (2018)			✓	✓̲		
Stuart-Hamilton et al. (2006)		✓	✓̲		✓	
Svedholm & Lindeman (2013)				✓̲	✓	
Tobacyk (1983)		✓			✓	✓̲
Tobacyk (1984)			✓̲		✓	
Van Elk (2013)	✓̲				✓	
Van Elk (2015)	✓̲				✓	
Van Elk (2017)	✓̲	✓			✓	
Wain & Spinella (2007)					✓̲	

Wierzbicki (1985)		✓̲			✓	
Willard & Norenzayan (2013)	✓̲					
Wilson (2018)			✓̲			
Wilson & French (2006)					✓̲	

Note: ✓̲ *= original category,* ✓ *= alternate category*

Appendix C – Table showing summary statistics for studies included in the "perceptual and cognitive biases" section of the systematic literature review

Study	Sample Size (% women)	Age Range, *M* (SD)	Bias	Tests Used	Key Findings
Lesaffre et al. (2020)	419 (69.4)	18-47, 20.50 (3.07)	C	Novel event explanation questionnaire, mental dice task	<u>Ns.</u> corr. paranormal beliefs and repetition avoidance + corr. paranormal beliefs and confirmation bias (r = .42, p < .001) – higher paranormal belief scores associated with explaining a magic performance as psychic
Drinkwater et al. (2019)	174 (75.9)	18-62, 24.62(/)	P	PRI	Paranormal belief predicts risk perception (b = .059, t=3.902, p < .001)
Griffiths et al. (2019)	160 (96.2)	/, 19.00(/)	P	Causal judgement task	+ corr. paranormal belief and causal judgement ratings (r(151) = .22, p =.005, BF = 4.76)
Barberia et al. (2018)	106 (81.1%)	*intervention* /, 21.57 (3.48) *control* /, 20.83 (2.65)	C	Novel confirmatory bias education intervention	- effect of intervention on precognition scores (t(102) = -2.62, p = .005, d = -0.52) <u>Ns.</u> effect of intervention on global paranormal belief, witchcraft, TRB, spiritualism, ELF, extra-terrestrial life and

					actual visits, or superstition scores (*p*s > .26)
Prike et al. (2018)	259 (58.9)	18-81, 35.44 (11.94)	C	BADE, BACE, liberal acceptance scenarios, jumping to conclusions task	- corr. paranormal belief and BADE ($r = -.22, p < .001$), BACE ($r = -.22, p < .001$) and + corr. with liberal acceptance ($r = .35, p < .001$) Liberal acceptance predicted paranormal belief ($F(1, 219) = 9.03, p = .003, \eta_p^2 = 0.04$, *BF*10 = 12.91)
Caputo (2017)	30 (80.0)	20-26, 21.70 (1.23)	P	SFQ	<u>Ns.</u> corr. overall paranormal belief and strange face illusions, but + corr. paranormal beliefs and two SFQ items (#7, $r = -.46, p < .01$; #12, $r = -.37, p < .05$)
Van Elk (2017)	53 (77.4)	/, 40.10 (15.40)	P	Computerised card guessing game	<u>Ns.</u> relationship paranormal belief and illusion of control (*F*s < 1)
Van Elk (2015)	55 (69.1)	/, 43.40 (/)	P	Novel face/house categorization task	Paranormal belief predicted accuracy on categorisation trials with 70% visual noise ($\beta = 0.353, p = .011$)
Blanco et al. (2015)	64 (81.3)	18-26, 18.69 (1.45)	P	Contingency task	+ corr. paranormal belief and illusion of control ($r = .28, p < .05$)

					+ effect paranormal belief on illusory control (β = .28, p = .02) for noncontingent task
Irwin et al. (2014)	124 (79.0)	18-65, 26.44 (10.43)	C	Computerised beads task, JTC subscale of CBQ, JTC subscale of DACOBS	- corr. TPB subscale of RPBS and both the 85:15 beads task (r = -.22, p < .05), and the 60:40 beads task (r = -.26, p < .01) + corr. TPB and CBQ (r = .23, p < .05) + corr. NAP subscale and DACOBS (r = .23, p < .01) Ns. corr. NAP and either: 85:15 beads task, 60:40 beads task, CBQ Ns. corr. TPB and DACOBS
Simmonds-Moore (2014)	95 (51.6)	20-76, 45.00 (13.10)	P	Visual and auditory detection task, each including one ESP, two degraded stimuli and one random trial	Faster response latencies for first guesses in visual degraded stimuli condition for paranormal believers compared to sceptics (X^2 = 5.44, df = 2, p = .036) More misidentifications of visual degraded stimuli for believers compared to sceptics (X^2 = 7.01, df = 2, p = .013) Believers more confident in first guesses for both visual (X^2 = 6.1, df = 2, p = .024) and

					auditory (χ^2 = 6.2, df = 2, p = .022) degraded stimuli
Willard & Norenzayan (2013)	479 (77.0) 825 (66.0)	18-41, 20.5 (/) 18-81, 34.7 (/)	C	DS, IDAQ, EQ, teleology statements	+ corr. paranormal belief and dualism (r = .43, $p \leq .01$), teleology (r = .18, $p \leq .01$), and anthropomorphism (r = .36, $p \leq .01$) in sample 1 + corr. paranormal belief and dualism (r = .33, $p \leq .01$), teleology (r = .19, $p \leq .01$), anthropomorphism (r = .31, $p \leq .01$), and mentalizing (r = .12, $p \leq .01$) in sample 2 + relationship paranormal belief and anthropomorphism (β = .29), dualism (β = .38), and teleology (β = .12) in sample 1 + relationship paranormal belief and anthropomorphism (β = .28), dualism (β = .29), and teleology (β = .05) in sample 2
Riekki et al. (2013)	47 (55.3)	20-50, 31.00 (/)	P	Novel face detection task including artefact face pictures vs non-face pictures	Paranormal believers more false alarms for non-face pictures compared to sceptics ($F(1, 36)$ = 7.95, p = .008, η_p^2 = .181)

					Paranormal believers more hits for artefact face pictures compared to sceptics ($F(1, 36) = 9.99$, $p = .003$, $\eta_p^2 = .217$) Paranormal believers lower response criteria ($F(1, 36) = 11.02$, $p = .002$, $\eta_p^2 = .234$) and higher correct detections ($F(1, 36) = 6.01$, $p = .019$, $\eta_p^2 = 143$) compared to sceptics Paranormal believers rated artefact faces as more face-like ($F(1, 37) = 6.25$, $p = .017$, $\eta_p^2 = .145$) and emotive ($F(1, 37) = 4.70$, $p = .037$, $\eta_p^2 = .113$) compared to sceptics
Van Elk (2013)	67 (71.6)	/, 28.30 (/)	P	Point-light-walker displays	+ corr. paranormal belief score and response bias for stimuli with 12 ($r = .45$, $p < .001$), 24 ($r = .29$, $p < .05$), 48 ($r = .43$, $p < .001$), and 96 distractors ($r =.41$, $p < .001$) **<u>Ns.</u>** corr. paranormal belief score and stimuli with 192 ($p = .25$) or 384 ($p < .621$) distractors Difference between sceptics and believers most pronounced for stimuli with low to

					intermediate number of visual distractors ($F(5, 290) = 3.1, p < .01, \eta_p^2 = .05$) Main effect of group on perceptual sensitivity ($F(1, 58) = 9.5, p < .005, \eta_p^2 = .14$) with sceptics showing higher perceptual sensitivity compared to believers Main effect of group on bias for reporting illusory patterns ($F(1, 58) = 8.4, p < .005, \eta_p^2 = .13$) with sceptics showing reduced response bias compared to believers
Krummenacher et al. (2010)	40 (00.0)	21-39, *bl*28.7 (4.8), *sc*28.4 (4.5)	P	Lexical and facial decision tasks	Main effect of group on response criterion ($F(1, 35) = 11.12, p = .002$), with believers demonstrating lower response criteria compared to sceptics + corr. paranormal belief and schizotypy ($rho = .92, p < .000$)
Rudski (2004)	275 (61.5)	18-25, / (/)	P	Illusion of control questionnaire	Higher paranormal belief scores for participants demonstrating illusion of control ($t(269) = 2.31, p = .02$) Higher scores on the superstition ($t(269) = 4.07, p < .001$) and precognition ($t(269) =$

					3.34, $p = .001$) subscales of the RPBS for participants demonstrating illusion of control
Schienle et al. (1996)	42 (54.8)	18-29, 23.10 (/)	P	Telepathy experiment	Paranormal believers gave higher estimates of success than sceptics ($F(1, 38) = 18.6$, $p < .001$) Difference between paranormal believers and sceptics in judgement accuracy ($t(20) = 3.19$, $p = .003$), with believers overestimating the number of total hits and sceptics judging the hit rate accurately
Blackmore & Moore (1994)	30 (33.3)	/, / (/)	P	Visual detection task and false identification question	+ corr. paranormal belief and false identification question responses, with believers claiming to make more false identifications of people in their daily lives ($r = .606$, $p < .001$) Ns. corr. paranormal belief and response type in visual detection task

Gagné & McKelvie (1990)	53 (37.7)	/, / (/)	P	Signal detection task and questionnaire	**Ns.** difference paranormal believers and sceptics in either the behavioural task or questionnaire (*p*s > .05)

Note: / = information not reported, P = perceptual biases, C = cognitive biases, bl = believers, sc = sceptics, + = positive, - = negative, corr. = correlation, ***Ns.*** *= nonsignificant, ESP = extrasensory perception, BADE = bias against disconfirmatory evidence, BACE = bias against confirmatory evidence, TRB = traditional religious beliefs, ELF = extraordinary lifeforms, PRI = Personal Risk Inventory (Hockey et al., 2000), SFQ = Strange-Face Questionnaire (Caputo, 2015), IDAQ = Individual Differences in Anthropomorphism Quotient (Waytz et al., 2010), DS = Dualism Scale (Stanovich, 1989), EQ = Empathy Quotient (Baron-Cohen & Wheelwright, 2004).*

Appendix D – Table showing summary statistics for studies in the "reasoning" section of the systematic literature review

Study	Sample Size (% women)	Age Range and *M* (SD)	Tests Used	Key Significant Findings
Pérez-Navarro & Martínez-Guerra (2020)	180 (71.1)	18-49, 23.20 (5.60)	RTQ	+ corr. paranormal belief and schizotypy (r =.28, p < .001) + corr. paranormal belief and conditional reasoning on tasks with paranormal content (r = .24, p < .001) Ns. corr. paranormal belief and probabilistic reasoning
Denovan et al. (2018)	725 (73.0)	18-64, 25.50 (9.40)	Perception of randomness, conjunction fallacy, paranormal perception of randomness, and paranormal conjunction fallacy tasks	+ corr. paranormal belief and schizotypy (r = .32, p < .001) - corr. paranormal belief and perception of randomness (r = -.24, p < .001) and conjunction fallacy (r = -.14, p < .001) - corr. paranormal belief and paranormal perception of randomness (r = -.37, p < .001) and paranormal conjunction fallacy (r = -.45, p < .001)
Prike et al. (2017)	434 (/)	/, / (/)	Scenario judgements questionnaire	+ corr. between paranormal belief and single conjunction errors in Experiment 1A (r = .21, p < .001) and 1B (r = .31, p < .001)

				+ corr. paranormal belief and conjunction errors in Experiment 1A (*r* = .25, *p* < .001) and 1B (*r* = .33, *p* < .001)
Dagnall et al. (2016A)	254 (73.0)	18-71, 26.66 (9.81)	Perception of randomness, conjunction fallacy, paranormal perception of randomness, and paranormal conjunction fallacy tasks	+ corr. paranormal belief and schizotypy scores (*r* = .25, *p* < .01) - corr. paranormal belief and perception of randomness (*r* = -.25, *p* < .01), conjunction fallacy (*r* = -.17, *p* < .01), paranormal perception of randomness (*r* = -.38, *p* < .01), and paranormal conjunction fallacy (*r* = -.49, *p* < .01) - effect of traditional paranormal belief on perception of randomness (β = -.39, *p* < .001) and conjunction fallacy (β = -.25, *p* < .05) - effect of new age philosophy on perception of randomness (β = -.31, p < .001) + effect of the unusual experiences O-LIFE subscale on the traditional paranormal belief (β = .31, p < .001) and new age philosophy (β = .31, p < .001) **Ns.** effect of NAP on conjunction fallacy

Dagnall et al. (2016B)	233 (74.0)	18-65, 23.00 (8.41)	Perception of randomness, conjunction fallacy, paranormal perception of randomness, and paranormal conjunction fallacy tasks	- corr. paranormal belief and perception of randomness (r = -.25, p < .01) and conjunction fallacy (r = -.12, p < .05) - corr. paranormal belief and paranormal perception of randomness (r = -.43, p < .01) and paranormal conjunction fallacy (r = -.46, p < .01)
Rogers et al. (2016)	213 (42.2)	18-74, 29.40 (10.40)	Scenario Judgements Questionnaire	+ corr. paranormal belief subscales and conjunction errors for paranormal and non-paranormal scenarios with confirmatory outcomes: ESP (r = .25, p < .001), PK (r = .30, p < .001), and LAD (r = .21, p < .01) subscales + corr. paranormal belief subscales and conjunction error rates for non-paranormal scenarios with confirmatory outcomes: ESP (r = .21, p < .01), PK (r = .30, p < .001), LAD (r = .19, p < .01) subscales + corr. paranormal scenarios with confirmatory outcomes: ESP (r = .18, p < .05), PK (r = .19, p < .01), LAD (r = .15, p < .05) subscales

				Paranormal belief predicted number of conjunction errors: ESP ($\beta = .18$, $p = .001$), PK ($\beta = .19$, $p < .001$), LAD ($\beta = .14$, $p = .004$) subscales
Dagnall et al. (2014)	305 (79.0)	18-65, 22.97 (8.19)	Perception of randomness, conjunction fallacy, paranormal conjunction fallacy, probability, and base rate tasks	- corr. overall probabilistic reasoning and paranormal belief on three scales - ASGS ($r = -.19$, $p < .01$), MMU-N ($r = -.23$, $p < .01$), and RPBS ($r = -.17$, $p < .01$) Perception of randomness predicted paranormal belief for the ASGS ($\beta = -.17$, $p = .04$), MMU-N ($\beta = -.16$, $p = .005$), and RPBS ($\beta = -.16$, $p = .008$) - corr. base rate test scores and ASGS ($r = -.12$, $p < .05$), MMU-N ($r = -.13$, $p < .05$), and RPBS ($r = -.07$, $p < .05$) - corr. probability test scores and ASGS ($r = -.14$, $p < .01$), MMU-N ($r = -.14$, $p < .01$), and RPBS ($r = -.10$, $p < .05$) - corr. perception of randomness and ASGS ($r = -.18$, $p < .01$), MMU-N ($r = -.19$, $p < .01$), and RPBS ($r = -.17$, $p < .01$)

				Ns. corr. paranormal belief and conjunction fallacy for ASGS, MMU-N and RPBS
Rogers et al. (2009)	200 (59.0)	18-56, 22.20 (5.30)	Scenario judgements questionnaire	Paranormal believers made more conjunction errors overall compared to sceptics ($F(1, 195) = 53.89$, $p < .001$) Sceptics made fewer conjunction errors for paranormal events compared to non-paranormal events ($F(1, 96) = 32.73$, $p < .001$) Believers made fewer conjunction errors for paranormal events compared to non-paranormal events ($F(1, 100) = 22.19$, $p < .001$) Believers made more conjunction errors compared to sceptics for both the probability ($F(1, 97) = 43.77$, $p < .001$) and frequency ($F(1, 97) = 16.10$, $p < .001$) questions
Dagnall et al. (2007)	96 (77.1)	18-47, 20.39 (4.85)	Perception of randomness, conjunction fallacy, base rate, and probability tasks	- corr. paranormal belief and overall number of correct responses ($r = -.24$, $p = .018$) Perception of randomness scores predicted paranormal belief ($\beta = -.32$, $p = .002$) - corr. perception of randomness and paranormal belief ($r = -.32$, $p = .001$)

				Participants low in paranormal belief solved more perception of randomness questions than those high in paranormal belief ($t(94) = 2.50$, $p = .014$) Ns. difference between high and low believers for base rate, probability, or conjunction fallacy questions ($p > .05$)
Lawrence & Peters (2004)	174 (27.6)	18-86, 54.0 (/)	DR task	Difference in number of overall DR errors ($U = 2691.5$, $p = .002$) with strong believers making more errors than weak believers Strong believers made more errors on both the control ($U = 2925.5$, $p = .01$) and antiparanormal ($U = 2668.5$, $p = .001$) statements Ns. difference in number of errors for proparanormal statements between strong and weak believers ($p = .10$)
Musch & Ehrenberg (2002)	123 (56.1)	/, 24.70 (5.60)	Probabilistic test battery	+ corr. paranormal belief and probabilistic reasoning errors ($r = .22$, $p < .05$) + corr. paranormal belief and lower cognitive ability ($r = .50$, $p < .01$) + corr. paranormal belief and lower cognitive ability when controlling for probabilistic reasoning skills ($r = .47$, $p < .01$)

Bressan (2002)	111 (65.8) 103 (48.5)	19-62, / (/) 19-61, / (/)	Probabilistic reasoning questionnaire	- corr. paranormal belief and probabilistic tasks for representativeness bias to sample size (*rho* = -.38, $p <$ .001), and representativeness applied to random sequences (*rho* = -.27, p = .006) in sample two **Ns.** difference in probabilistic reasoning scores between believers and sceptics in sample one
Roberts & Seager (1999)	65 (63.1)	18-75, 33.70 (11.30)	Probabilistic reasoning and conditional reasoning tasks	- corr. paranormal belief and reasoning ability (r = -.25, $p <$ .05) **Ns.** corr. paranormal belief and probabilistic reasoning ($p >$ .05) - corr. overall conditional reasoning score and paranormal belief (r = -.27, $p <$ .05)
Blackmore (1997)	6238 (59.0)	/, / (/)	Probabilistic reasoning questionnaire	Believers claimed more of the statements were true for them (M = 2.56) compared to sceptics (M = 2.23), and these differences were significant for 5 out of 10 items **Ns.** difference in probability misjudgements between believers and sceptics
Brugger et al. (1991)	95 (/)	/, / (/)	Probability judgement task	Believers demonstrated higher illusion of control and probabilistic reasoning errors compared to sceptics (χ^2 = 3.92, p = .023)

Brugger et al. (1990)	48 (50.0)	20-40, / (/)	Mental dice task	Believers made fewer repetitions compared to sceptics (t = 1.8, $p < .05$), and made increased errors when making probability judgements (t = 2.8, $p < .005$)
Wierzbicki (1985)	64 (53.1%)	/, / (/)	Conditional reasoning statements	+ relationship between paranormal belief scores and reasoning errors ($F(1, 63) = 6.37$, $p < .05$) + corr. paranormal belief scores and total number of reasoning errors (r = .28, $p < .05$) Ns. interaction between paranormal belief and logical form Ns. interaction between paranormal belief and content (symbolic and parapsychological problems)

*Note: / = information not reported, + = positive, - = negative, corr. = correlation, **Ns.** = nonsignificant, ESP = extrasensory perception, PK = psychokinesis, LAD = life after death, NAP = new age philosophy, DR = deductive reasoning, RTQ = Reasoning Task Questionnaire (Blackmore & Troscianko, 1985), ASGS = Australian Sheep-Goat Scale (Thalbourne & Delin, 1993), RPBS = Revised Paranormal Belief Scale (Tobacyk, 2004), MMU-N = Manchester Metropolitan University New (Dagnall et al., 2010)*

Appendix E – Table showing summary statistics for studies included in the "intelligence, critical thinking, and academic performance" section of the systematic literature review

Study	Sample Size (% women)	Age Range and *M* (SD)	Focus of Study	Tests Used	Key Significant Findings
Betsch et al. (2020)	599 (60.0)	18-81, 33.63 (11.38)	I	IQ	Paranormal beliefs negatively predicted by IQ when sex is excluded from the model (β = -.158, $p < .001$)
Andrews & Tyson (2019)	687 (54.4)	18-65, 24.62 (/)	C	Average university assignment grade	- corr. mean grade and paranormal beliefs ($r(162)$ = -0.388, p = .001), higher paranormal beliefs 'soft science' and 'artistic' students than 'hard science' students
Wilson (2018)	340 (52.7)	/, / (/)	C	Science and critical thinking course	Belief in psychics (T_{309} = 3.14, $p < .001$), witchcraft (T_{308} = 5.68, $p < .001$), spiritualism (T_{309}) = 9.23, $p < .001$), monsters (T_{309}) = 9.13, $p < .001$), precognition (T_{308} = 8.53, $p < .001$), aliens (T_{310} = 13.49, $p < .001$), and alternative medicine (T_{310} = 15.83, $p < .001$) lowered following critical thinking course **Ns.** difference in superstition following critical thinking course
McLean & Miller (2010)	47 (70.2)	/, 21.33 (1.74)	C	ATS, WGCTA-S, course in critical thinking skills	Paranormal belief decreased following critical thinking course ($F(1, 44)$ = 48.71, $p < .001$)

					Main effect of time, lower levels of paranormal belief post-test compared to pre-test ($F(1, 44) = 44.63$, $p < .001$)
Stuart-Hamilton et al. (2006)	73 (/)	60-84, 71.12 (5.21)	I	RPM, MHVT, and probability tests	Ns. corr. paranormal belief and intelligence (fluid or crystallised), paranormal belief and probability tests
Hergovich & Arendasy (2005)	180 (59.4)	18-37, 24.54 (3.61)	C, I	CCTT, WGCTA, WMT, RPM Rasch Model	- corr. reasoning ability and traditional paranormal belief ($r = -.22$, $p < .01$), superstition ($r = -.23$, $p < .01$), and traditional religiosity ($r = -.23$, $p < .01$) Univariate effect of reasoning ability for traditional paranormal belief ($F(1, 166) = 8.62$, $p < .01$) and new age philosophy ($F(1, 166) = 6.35$, $p < .05$) Ns. effect of critical thinking ability on paranormal belief ($p > .05$)
Roe (1999)	117 (65.8)	/, 21.00 (/)	C	Evaluation of an experimental report	Ns. main effects for paranormal belief and critical thinking ability
Morgan & Morgan (1998)	124 (65.0)	18-54, 24.30 (/)	C	WGCT	- corr. Superstitious Belief subscale and overall WGCT score ($r = -.19$, $p < .05$), as well as with the Inference subscale of the WGCT ($r = -.17$, $p < .05$) - corr. Traditional Religious Belief subscale and Evaluation of Arguments subscale of the WCGT ($r = -.18$, $p < .05$)

					- corr. Spiritualism subscale and the Inference (*r* = -.21, *p* < .05) and Recognition of Assumptions (*r* = -.18, *p* < .05) subscales of the WGCT Ns. corr. full scale scores for paranormal belief and critical thinking
Smith et al. (1998)	60 (40.0)	18-37, 21.10 (/)	I	APM Set 1	- corr. overall paranormal belief and intelligence (*rho* = -.29, *p* < .01) - corr. intelligence and spiritualism (*rho* = -.35, *p* < .01), psi (*rho* = -.30, *p* < .01), and precognition (*rho* = -.25, *p* < .05)
Royalty (1995)	97 (49.5)	18-41, 20.5 (/) 18-81, 34.7 (/)	C	CCTT Level Z, WAIS-IS	Ns. corr. paranormal belief and general critical thinking ability, or IQ
Tobacyk (1984)	307 (45.0)	*m* /, 19.70 (1.70) *f* /, 19.3 (1.7)	C	GPA	- corr. GPA and Witchcraft (*r* = -.13, *p* < .03) and Superstition (*r* = -.20, *p* < .001) subscales Ns. corr. total paranormal belief and GPA
Alcock & Otis (1980)	26 (/)	/, / (/)	C	WGCTA	Sceptics demonstrated a higher level of critical thinking ability than believers (*t*(24) = 2.07, *p* < .05)

Note: / = information not reported, C = cognitive ability, I = intelligence, m = males, f = females, + = positive, - = negative, corr. = correlation, Ns. = nonsignificant, ATS = Assessment of Thinking Skills (Wesp & Montgomery, 1998), WGCTA-S = Watson-Glaser Critical Thinking Appraisal Form S (Watson & Glaser, 1994), WGCTA = Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 2002; Watson & Glaser, 1980; Watson & Glaser, 1964), RPM = Raven's Progressive Matrices (Raven et *al., 2000), RPM Rasch Model = Raven's Progressive Matrices Rasch Model (Rasch, 1960), MHVT = Mill Hill Vocabulary Test (Raven et al., 1998), CCTT = Cornell Critical Thinking Test (Ennis & Millman, 1985), WMT = Wiener Matrizen Test (Formann & Piswanger, 1979), APM = Advanced Progressive Matrices (Raven, 1976), WAIS-IS = Wechsler Adult Intelligence Scale Information Subtest (Wechsler, 1955), GPA = Grade Point Average*

Appendix F – Table showing summary statistics for studies included in the "thinking style" section of the systematic literature review

Study	Sample Size (% women)	Age Range and *M* (SD)	Tests Used	Key Significant Findings
Ballová Mikušková & Čavojová (2020)	473 (50.5) 492 (53.0)	18-67, 41.54 (13.75) 18-66, 39.79 (13.68)	CRT	- corr. paranormal beliefs and analytical thinking ($r = -.149$, $p < .01$) in study 1 <u>Ns.</u> corr. paranormal beliefs and analytical thinking in study 2
Rizeq et al. (2020)	321 (81.9)	18-30, 19.36 (2.09)	AOT, CRT, verbal and non-verbal reasoning tasks	- corr. paranormal belief and cognitive ability ($r = -.14$, $p < .05$), as well as scores on the CRT ($r = -.22$, $p < .05$), and AOT ($r = -.44$, $p < .05$) AOT scored predicted paranormal belief scores ($\beta = -.39$, $p < .001$) <u>Ns.</u> cognitive ability and cognitive reflection did not predict paranormal belief
Branković (2019)	257 (43.0)	/, 21.94 (5.74)	REI	Intuitive cognitive style predicted belief in ESP ($\beta = .35$, $p < .001$) Traditional superstitious beliefs best predicted by a lack of rational engagement ($\beta = -.27$, $p = .001$) and self-rated intuitive ability ($\beta = .20$, $p = .018$)

Rogers et al. (2019)	343 (61.5)	18-80, 40.50 (12.60)	REI	Partial correlation between intuitive thinking style and paranormal belief when controlling for sampling method ($r = .57, p < .001$) **Ns.** corr. paranormal belief and analytical thinking
Ståhl & van Prooijen (2018)	343 (62.0) 322 (47.0)	/, 35.41 (12.01) /, 34.95 (10.86)	CRT, CRT-2, Numeracy test, WST	Analytic cognitive style associated with weaker paranormal belief ($b = -0.44, SE = 0.09, t = -4.93, p < .001$) in sample one Analytic cognitive style associated with weaker paranormal belief ($b = -0.18, SE = 0.05, t = -3.81, p < .001$) in sample two Cognitive ability contributes to scepticism when controlling for analytic cognitive style for individuals who value epistemic rationality ($b = -0.38, SE = 0.08, t = -4.90, p < .001$) **Ns.** effect analytic cognitive style on paranormal belief when controlling for cognitive ability
Lindeman & Svedholm-Häkkinen (2016)	258 (63.6)	18-65, 31.81 (9.89)	Mental rotation test, CRT, RI, maths and physics school grades	- corr. paranormal belief and analytical thinking ability ($r = -.30, p < .001$), analytical thinking style ($r = -.33, p < .001$), mental rotation ($r = -.13, p < .05$), and grades in physics ($r = -.17, p < .01$) and maths ($r = -.16, p < .01$) + corr. paranormal belief and intuitive thinking style ($r = .50, p < .001$) Intuitive thinking predicted paranormal beliefs ($\beta = .34, p < .001$)

Lasikiewicz (2016)	82 (77.0)	18-62, 29.96 (12.53)	REI	+ corr. paranormal belief an intuitive thinking (*r* = .31, $p < .01$) Paranormal belief predicted by both analytic (β = .246, *p* = .030) and intuitive thinking styles (β = .294, *p* = .018)
Irwin (2015)	94 (58.5)	18-73, 34.73 (15.19)	REI	+ corr. paranormal belief and intuitive thinking (*rho* = .41, $p <$ .05) Intuitive thinking style predicted paranormal belief (*beta* = .39, $p < .001$) Relationship between paranormal belief and intuitive thinking style ($F(3, 90) = 6.65$, $p < .001$) <u>Ns.</u> relationship paranormal belief and analytical thinking style (*p* = .336)
Majima (2015)	246 (76.4)	18-81, 25.00 (13.00)	IPSI-SF, logical reasoning task	Intuitive thinking style (β = 0.24, $p < .001$) and analytical thinking style (β = 0.14, *p* = .039) both predicted paranormal belief <u>Ns.</u> relationship paranormal belief and cognitive ability (*p* = .956)
Svedholm & Lindeman (2013)	50 (74.0) 458 (77.1)	19-62, 34.00 (/) 18-65, 27.00 (7.90)	FIS, NFC, AOT, AET, REI	+ corr. paranormal belief and intuitive thinking (*r* = .37, *p* = .01) in sample one + corr. paranormal belief and inhibition errors (*r* = .22, *p* = .14) in sample one

				Paranormal belief predicted by intuitive thinking in sample one (β = .368, p = .009) and sample two (β = .460, p < .001) - corr. paranormal belief and AOT scores in sample one (r = -.19, p = .19) and sample two (r = -.41, p < .001) - corr. paranormal belief and NFC scores in sample one (r = -.20, p = .16) and sample two (r = -.21, p < .001) - corr. paranormal belief and AET scores (r = -.30, p = .04) in sample one
Genovese (2005)	96 (71.1)	20-57, 28.00 (/)	10-item REI	+ corr. paranormal belief and intuitive thinking (r = .32, p < .01) + corr. paranormal belief and the cognitive perceptual (r = .65, p < .01) and disorganized (r = .27, p < .01) subscales of the SPQ-B **<u>Ns.</u>** corr. paranormal belief and rational thinking, or between paranormal belief and interpersonal subscale of SPQ-B
Hergovich (2003)	91 (57.1) 150 (57.3)	18-60, 34.15 (13.98) /, 37.28 (13.31)	GWT, EFT	- corr. field dependence and paranormal belief (r = -.39, p < .001) in study one - corr. field dependence and superstition subscale of PBS (r = .45, p < .01)

				Ns. corr. field dependence and total paranormal belief in study two Ns. main effect of field dependence on paranormal belief in study three
Gianotti et al. (2001)	24 (54.2)	/, / (/)	Novel word association task	Interaction effect between belief group and stimulus type ($F(1, 22) = 6.92$, $p < .015$) Believers produced more rare associations compared to sceptics for unrelated word pairs (19.3% vs 12.0%, $p < .04$) Ns. differences in response latencies for believers and sceptics ($p = .087$)

*Note: / = information not reported, + = positive, - = negative, corr. = correlation, **Ns.** = nonsignificant, AOT = Actively Open-Minded Thinking Scale (Stanovich et al., 2016; Stanovich, 1999), CRT = Cognitive Reflection Test (Frederick, 2005), CRT-2 = Cognitive Reflection Test-2 (Thompson & Oppenheimer, 2016), REI = Rational-Experiential Inventory (Pacini & Epstein, 1999), WST = WordSum Test (Huang & Hauser, 1998), RI = Rational/Experiential Inventory (Norris & Epstein, 2011), IPSI-SF = Information-Processing Style Inventory Short Form (Naito et al., 2004), FIS = Faith in Intuition Scale (Pacini & Epstein 1999), NFC = Need for Cognition scale (Cacioppo et al., 1984), AET = Argument Evaluation Test (Stanovich & West, 1997), 10-Item REI = 10-Item Rational-Experiential Inventory (Epstein et al., 1996), GWT = Gestaltwahrnehmungs Test (Hergovich & Hörndler, 1994), EFT = Embedded Figures Test (Witkin et al., 1971)*

Appendix G – Table showing summary statistics for studies included in the "executive function and memory" section of the systematic literature review

Study	Sample Size (% women)	Age Range and *M* (SD)	Focus of Study	Tests Used	Key Significant Findings
Gray & Gallo (2016)	84 (53.6) 115 (48.7) 95 (48.4)	*bl* /, 27.3 (/) *sc* /, 26.5 (/) *bl* /, 27.04 (/) *sc* /, 27.44 (/) *bl* /, 27.4 (/) *sc* /, 27.0 (/)	M	DRM task, CRT, IIT, RSPAN, OSPAN, SILS, AET, RAT	*Study one:* Sceptics remembered more studied words in DRM task compared to believers for both no warning ($t(82) = 3.23, p < .001$) and warning conditions ($t(82) = 2.88, p = .005$) Sceptics better than believers at identifying critical lure as a missing item in DRM task ($t(82) = 2.54, p = .01$) Sceptics recalled fewer critical lures than believers for the warning condition in DRM task ($t(82) = 2.50, p = .01$) Sceptics solved more logic problems on the SILS ($t(81) = 2.56, p = .01$), and identified more words in the vocabulary test of the SILS ($t(81) = 3.17, p < .01$) than believers **Ns.** difference in number of false recalls of non-critical words ($p = .41$) or of critical lures ($p = .74$) recalled in no warning condition of the DRM task between sceptics and believers

Ns. difference in number of falsely recalled noncritical words between sceptics and believers for warning condition of DRM task ($p = .24$)

Ns. difference between believers and sceptics on either the red-word test ($p = .89$), picture test ($p = .43$) or exclusion test ($p = .15$) of the CRT

Study two:

Believers remembered more words per list in correct serial position compared to sceptics for RSPAN ($t(113) = -2.21$, $p = .03$)

Ns. difference between believers and sceptics for OSPAN ($p = .16$), RAT ($p = .06$) or AET ($p = .24$)

Study three:

Sceptics better than believers at identifying critical lure as missing item in the DRM task ($t(93) = 3.68$, $p < .001$)

Ns. difference between sceptics and believers for recall of studied words ($p = .31$), false recall of critical lures ($p = .28$) or false recall of non-critical words ($p = .61$) in DRM task

					Ns. difference between sceptics and believers for RSPAN (p = .13), OSPAN (p = .66), or for the logic (p = .10) and vocabulary (p = .09) tests of the SILS *Studies one and three – pooled data:* Sceptics solved more of the SILS logic problems ($t(168) = 3.03, p < .01$), and identified more words in the vocabulary SILS test ($t(168) = 3.42, p < .01$) than believers *Studies two and three – pooled data:* Believers remembered more words per list in correct serial position for RSPAN task ($t(208) = -2.67, p = .008$) than sceptics Ns. difference between believers and sceptics for OSPAN (p = .19), RAT (p = .07) or AET (p = .06)
Lindeman et al. (2011)	26 (61.5)	*bl* /, 34.6 (/) *sc* /, 32.2 (/)	EF	Stroop task, WCST	In total, believers and sceptics performed differently on all four subscales of the WCST ($F(5, 20) = 3.47, p = .02, \eta^2 = .398$) Believers' had higher total errors ($p < .01$), non-perseverative errors ($p < .01$), perseverative errors ($p < .03$), and lower categories correct ($p < .05$) compared to sceptics Ns. difference between believers and sceptics for Stroop

Wain & Spinella (2007)	213 (66.2)	18-83, 28.0 (11.9)	EF	EFI	- corr. total paranormal beliefs and: overall EF ($r = -.19, p < .01$), impulse control ($r = -.29, p < .001$), and organisation ($r = -.23, p < .001$) - corr. superstition subscale and: overall EF ($r = -.30, p < .001$), motivational drive ($r = -.17, p < .05$), empathy ($r = -.23, p < .01$), and organisation ($r = -.33, p < .01$) - corr. ANP subscale and: overall EF ($r = -.19, p < .01$), impulse control ($r = -.32, p < .001$), and organisation ($r = -.26, p < .001$) - corr. psychic beliefs subscale and: impulse control ($r = -.25, p < .001$) and organisation ($r = -.13, p < .05$) - corr. witchcraft subscale and impulse control ($r = -.22, p < .01$) - corr. TRB subscale and: motivational drive ($r = .14, p < .05$) and empathy ($r = -.21, p < .01$) **Ns.** corr. total paranormal beliefs and: motivational drive, empathy or strategic planning **Ns.** corr. ANP subscale and either motivational drive, empathy, or strategic planning

					Ns. corr. psychic beliefs subscale and either motivational drive, empathy, strategic planning, or overall EF Ns. corr. superstition subscale and either impulse control or strategic planning Ns. corr. witchcraft subscale and either motivational drive, empathy, organisation, strategic planning, or overall EF Ns. corr. TRB subscale and either impulse control, organisation, strategic planning, or overall EF
Wilson & French (2006)	100 (58.0)	23-52, 33.4 (9.87)	M	NCQ	Higher ASGS scores ($t(98) = 3.49$, $p = .001$) and higher scores on the belief subscale of the AEI ($t(98) = 4.26$, $p < .001$) for participants reporting false memories compared to participants not reporting false memories Paranormal beliefs measured with belief subscale of AEI predicted false memory responses ($\beta = .28$, $p = .04$) Ns. ASGS scores did not predict false memory responses
Greening (2002) †	16 (81.3) 52 (78.8) 53 (79.2)	/, 22.4 (4.83) /, 25.4 (7.66) /, 56.3 (8.55)	M	False memories questionnaire	+ corr. paranormal belief and number of false memories in study 2.1. ($r = .52$, $p < .05$) Ns. corr. paranormal beliefs and number of false memories in studies 2.2 ($p = .17$) and 2.3 ($p = .62$)

Dudley (1999)	56 (64.3)	18-24, / (/)	M	Digit span test	Higher paranormal belief scores in experimental group (working memory restricted) compared to control group ($F(1,50) = 5.44$, $p < .05$)

Note: / = information not reported, M = memory, EF = executive function, bl = believers, sc = sceptics, **+** *= positive, - = negative, corr. = correlation,* ***Ns.*** *= nonsignificant, DRM = Deese-Roediger-McDermott (Roediger & McDermott, 1995), CRT = Criterial Recollection Task (Gallo, 2013), IIT = Imagination Inflation Task (Garry et al., 1996), RSPAN = Reading-Span Task (Daneman & Carpenter, 1980), OSPAN = Operation Span Task (Turner & Engle, 1989), SILS = Shipley Institute of Living Scale (Zachary, 1986), AET = Argument Evaluation Task (Stanovich & West, 1997), RAT = Remote Associations Test (Mednick, 1962), WCST = Wisconsin Card Sorting Test (Berg, 1948; Grant & Berg, 1948), EFI = Executive Function Index (Spinella, 2005), ANP = anomalous natural phenomena, TRB = traditional religious beliefs, NCQ = News Coverage Questionnaire (Wilson & French, 2006), ASGS = Australian Sheep-Goat Scale (Thalbourne 1995; Thalbourne & Delin, 1993), AEI = Anomalous Experiences Inventory (Kumar et al., 1994)*

Appendix H – Table showing summary statistics for studies included in the "other cognitive functions" section of the systematic literature review

Study	Sample Size (% women)	Age Range and *M* (SD)	Focus of Study	Tests Used	Key Significant Findings
Palmer et al. (2007)	40 (00.0)	20-40, / (/)	ISL	Arrow prediction task	Difference in schizotypy scores between believers and sceptics ($t(33) = 15.1, p < .001$) with believers scoring higher than sceptics <u>Ns.</u> difference in sequence learning scores between believers and sceptics
Pizzagalli et al. (2001)	24 (100.0)	*bl* /, 26.3 (6.2) *sc* /, 26.8 (4.3)	ISP	Semantic priming task	<u>*Word types (prime-target relation):*</u> Group x VF interaction for indirectly related targets ($F(1, 22) = 6.32, p < .02$) Main effect of group x VF x category ($F(2, 44) = 5.11, p , < .01$) Believers had shorter reaction times for indirectly related target words presented in the LVF compared to sceptics ($p < .001$) Sceptics shorter reaction times for directly related target words presented in the LVF compared to both indirectly related (900ms ± 90 vs 1081ms ± 209, $p < .001$) and unrelated target words (900ms ± 90 vs 1058ms ±)

Sceptics shorter reaction times for directly related target words presented in the RVF compared to both indirectly related (846ms ± 109 vs 924 ± 160, $p < .05$) and unrelated target words (846ms ± 109 vs 1004 ± 177)

Sceptics' reaction times to indirectly related targets differed from those to unrelated targets only in the RVF

Believers had shorter reaction times in the LVF for directly related target words compared to unrelated target words (864ms ± 146 vs 1014ms ± 156), and for indirectly related words compared to unrelated target words (912ms ± 177 vs 1014ms ± 156)

Believers had shorter reaction times in the RVF for directly related target words compared to unrelated target words (766ms ± 114 vs 961ms ± 271), and for indirectly related words compared to unrelated target words (875ms ± 199 vs 961ms ± 271).

Believers' reaction times to directly related targets differed to indirectly related targets only in the RVF ($p < .005$)

Ns. difference in reaction times for indirectly related target words presented in the RVF between believers and sceptics ($p > .03$)

					<u>**Ns.**</u> difference in believers' reaction times to directly and indirectly related target words in the LVF <u>**Ns.**</u> group x VF interaction for directly related or unrelated targets *<u>Type of priming:</u>* Group x VF x type of priming interaction ($F(1, 22) = 10.74$, $p < .005$) For sceptics, direct semantic priming differed from indirect sematic priming in both the LVF (158ms ± 214 vs -23ms ± 92, $p < .001$) and the RVF (158ms ± 112 vs 80ms ± 80, $p < .05$) – shorter reaction times for indirect semantic priming For believers, direct semantic priming differed from indirect semantic priming in the RVF (196ms ± 180 vs 87ms ± 100, $p < .001$) – shorter reaction times for indirect semantic priming <u>**Ns.**</u> difference in reaction times between direct and indirect semantic priming in the LVF for believers
Irwin & Green (1998-99)	194 (57.2)	18-46, 22.2 (6.05)	CME	Central monitoring computer game	+ corr. paranormal belief and cognitive perceptual ($r = .35$, $p < .001$) and disorganised ($r = .24$, $p < .001$) subscales of SPQ-B <u>**Ns.**</u> corr. paranormal beliefs and interpersonal subscale of SPQ-B

					Ns. corr. paranormal beliefs and central monitoring performance ($p > .15$)
Tobacyk (1983)	110 (35.5)	*f* /, 20.6 (3.0) *m* /, 20.1 (1.7)	CC	RCRG	**Ns.** corr. paranormal beliefs and CC

Note: / = information not reported, bl = believers, sc = sceptics, f = females, m = males, ISL = implicit sequence learning, ISP = implicit semantic priming, VF = visual field, LVF = left visual field, RVF = right visual field, CME = central monitoring efficiency, RE = reasoning errors, CC = cognitive complexity, **+** *= positive, - = negative, corr. = correlation,* ***Ns.*** *= nonsignificant, SPQ-B = Schizotypal Personality Questionnaire Brief (Raine & Benishay, 1995), RCRG = Role Construct Repertory Grid (Kelly, 1955)*

Appendix I – Table showing measures of paranormal beliefs used in the 71 studies included in the systematic literature review

RPBS	Novel	ASGS	PBS	Rasch RPBS	Okkulten Glaubens	BPS-O	BPS*	MM U-N	MM U-PS	SSUB	OS	PS	AEI
Morgan & Morgan (1998)	Alcock & Otis (1980)†	Roe (1999)	Tobacyk (1983)	Wain & Spinella (2007)	Schienle et al. (1996)	Wierzbicki (1985)	Genovese (2005)	Dagnall et al. (2014)	Drinkwater et al. (2019)	Irwin (2015)	Hergovich (2003)	Ståhl & van Prooijen (2018)	Rogers et al. (2019)
Irwin & Green (1998-99)	Brugger et al. (1990)	Wilson & French (2006)	Tobacyk (1984)	Dagnall et al. (2014)	Gianotti et al. (2001)	Roberts & Seager (1999)	Stuart-Hamilton et al. (2006)						
Dudley (1999)	Brugger et al. (1991)	Palmer et al. (2007)	Gagné & McKelvie (1990)	Irwin et al. (2014)	Pizzagalli et al. (2001)								
Rudski (2004)	Blackmore & Moore (1994)	Rogers et al. (2009)	Royalty (1995)	Dagnall et al. (2016A)									
Hergovich & Arendasy (2005)	Blackmore (1997)	Krummenacher at al. (2010)	Smith et al. (1998)	Dagnall et al. (2016B)									
Dagnall et al. (2007)	Bressan (2002)†	Dagnall et al. (2014)	Greening (2002)	Denovan et al. (2018)									
McLean & Miller (2010)	Musch & Ehrenberg (2002)†	Simmonds-Moore (2014)											
Lindeman et al. (2011)	Lawrence & Peters (2004)	Gray & Gallo (2016)											
Riekki et al. (2013)	Majima (2015)†	Rogers et al. (2016)											
Svedholm & Lindeman (2013)	Prike et al. (2017)†	Prike et al. (2018)											

Van Elk (2013) Willard & Norenzayan (2013) Blanco et al. (2015) Van Elk (2015) Lasikiewicz (2016) Lindeman & Svedholm-Häkkinen (2016) Caputo (2017) Van Elk (2017) Barberia et al. (2018)‡ Wilson (2018) Andrews & Tyson (2019) Mikušková & Cavojavá (2020)	Branković (2019)† Griffiths et al. (2019)† Bestch et al. (2020)† Rizeq et al. (2020)

Lesaffre et al. (2020)													
Pérez-Navarro & Martínez-Guerra (2020)													
24	14	10	6	6	3	2	2	1	1	1	1	1	1

*Note:† = papers that provided reliability statistics for their novel scales, ‡ = used a translated version of the original scale, * = Musch & Ehrenberg (2002) developed a novel scale that was later named the BPS and was used in two subsequent studies. RPBS = Revised Paranormal Belief Scale (Tobacyk 1988; 2004), ASGS = Australian Sheep-Goat Scale (Thalbourne & Delin, 1993), PBS = Paranormal Belief Scale (Tobacyk & Milford, 1982), Rasch RPBS = Rasch devised Revised Paranormal Belief Scale (Lange et al., 2000), BPS-O = Belief in the Paranormal Scale (Original; Jones et al., 1977), BPS = Belief in the Paranormal Scale (Musch & Ehrenberg, 2002), MMU-N = Manchester Metropolitan University New (see Dagnall et al., 2010), MMU-PS = Manchester Metropolitan University Paranormal Scale (see Dagnall et al., 2010), SSUB = Survery of Scientifically Unsubstantiated Beliefs (Irwin & Marks, 2013), OS = Occultism Scale (Böttinger, 1976), PS = Paranormal Scale (Orenstein, 2002), AEI = Anomalous Experiences Inventory (Gallagher et al., 1994; includes a 'belief' subscale).*

Appendix J – Tobacyk's (2004) Revised Paranormal Belief Scale (RPBS)

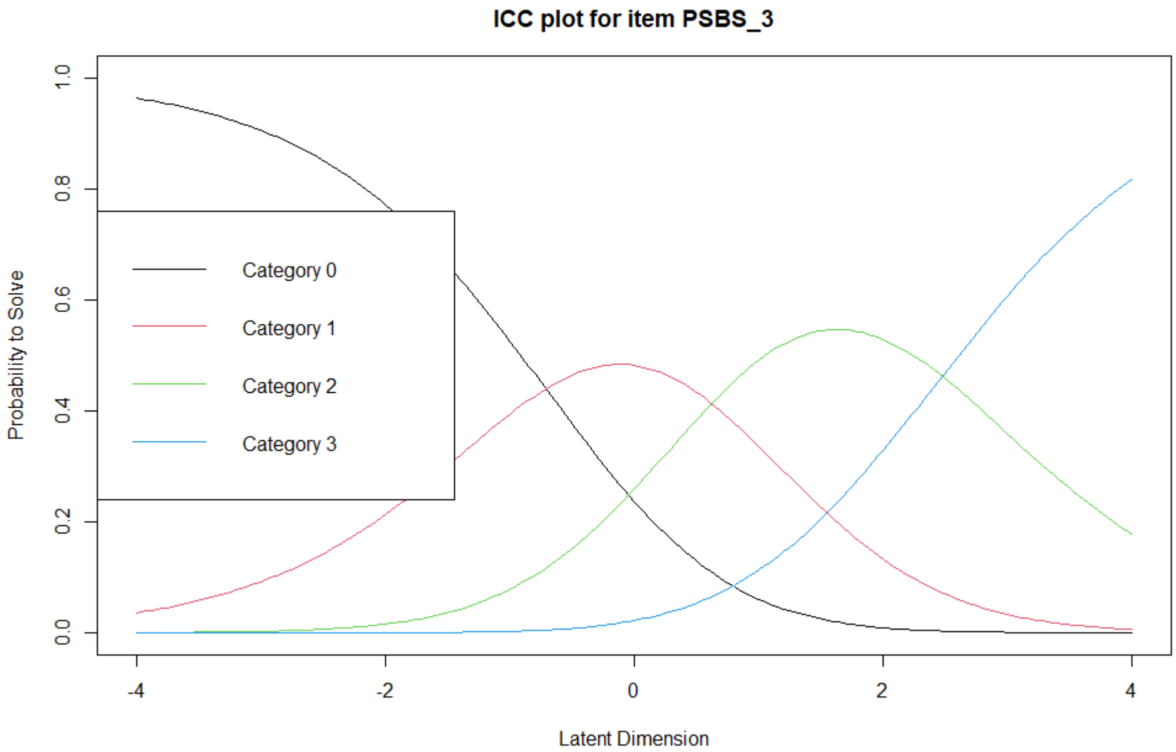
Please put a number next to each item to indicate how much you agree or disagree with that item. Use the numbers as indicated below. There are no right or wrong answers. This is a sample of your own beliefs and attitudes. Thank you.

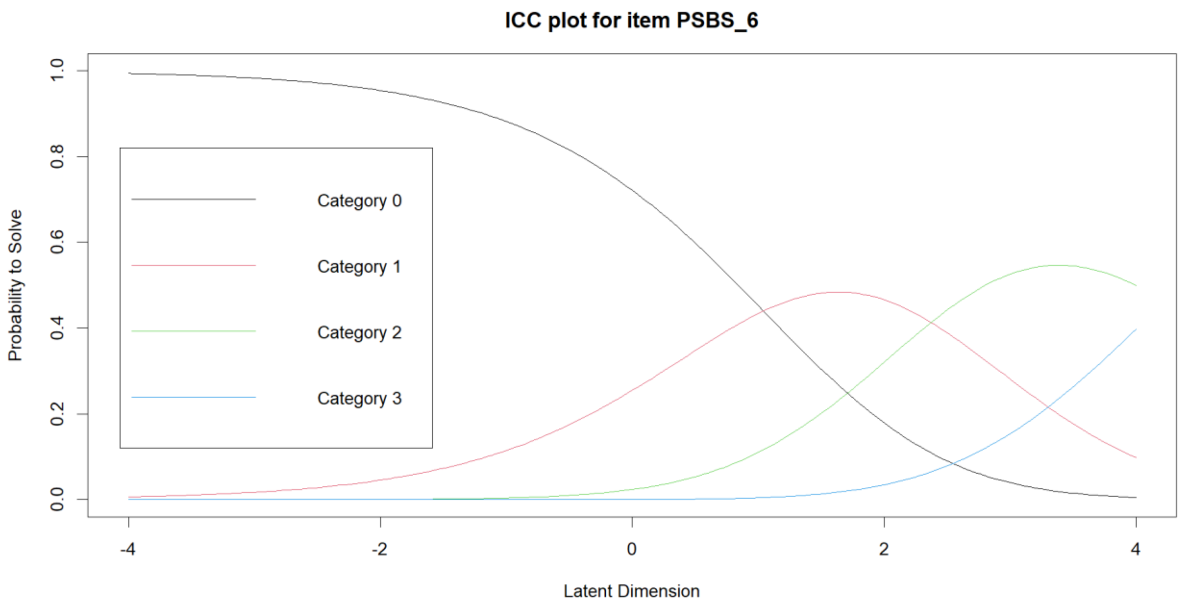
1=Strongly Disagree 2=Moderately Disagree 3=Slightly Disagree
4=Uncertain 5=Slightly Agree 6=Moderately Agree 7=Strongly Agree

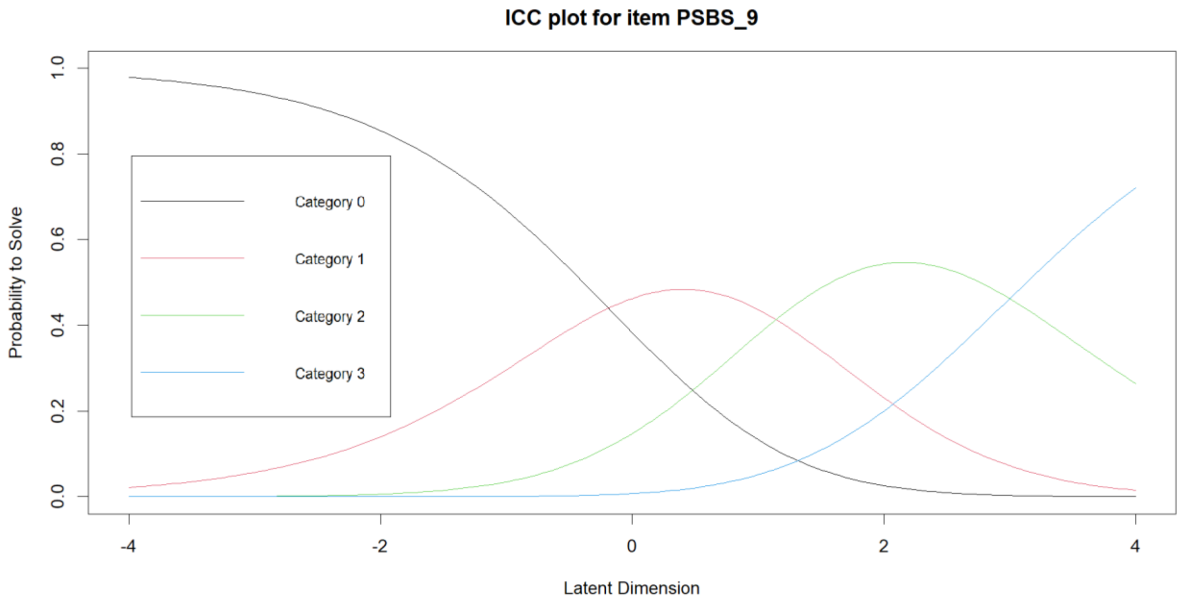
1. The soul continues to exist though the body may die.
2. Some individuals are able to levitate (lift) objects through mental forces.
3. Black magic really exists.
4. Black cats can bring bad luck.
5. Your mind or soul can leave your body and travel (astral projection).
6. The abominable snowman of Tibet exists.
7. Astrology is a way to accurately predict the future.
8. There is a devil.
9. Psychokinesis, the movement of objects through psychic powers, does exist.
10. Witches do exist.
11. If you break a mirror, you will have bad luck.
12. During altered states, such as sleep or trances, the spirit can leave the body.
13. The Loch Ness monster of Scotland exists.
14. The horoscope accurately tells a person's future.
15. I believe in God
16. A person's thoughts can influence the movement of a physical object.
17. Through the use of formulas and incantations, it is possible to cast spells on persons.
18. The number "13" is unlucky.
19. Reincarnation does occur.
20. There is life on other planets.
21. Some psychics can accurately predict the future.
22. There is a heaven and a hell.
23. Mind reading is not possible.
24. There are actual cases of witchcraft.
25. It is possible to communicate with the dead.
26. Some people have an unexplained ability to predict the future.

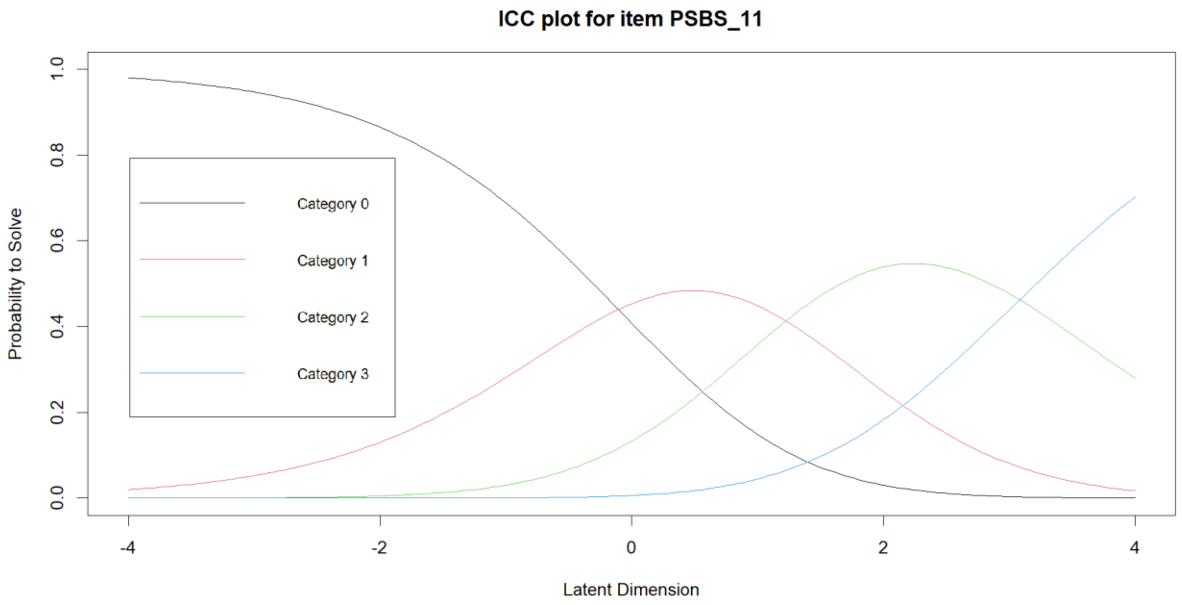
Note. Item 23 is reverse scored. Traditional Religious Belief = Mean of Items (1, 8, 15, 22);
Psi = Mean of Items (2, 9, 16, 23); Witchcraft = Mean of Items (3, 10, 17, 24);
Superstition = Mean of Items (4, 11, 18); Spiritualism = Mean of Items (5, 12, 19, 25)
Extraordinary Life Forms = Mean of Items (6, 13, 20); Precognition = Mean of Items (7, 14, 21, 26).

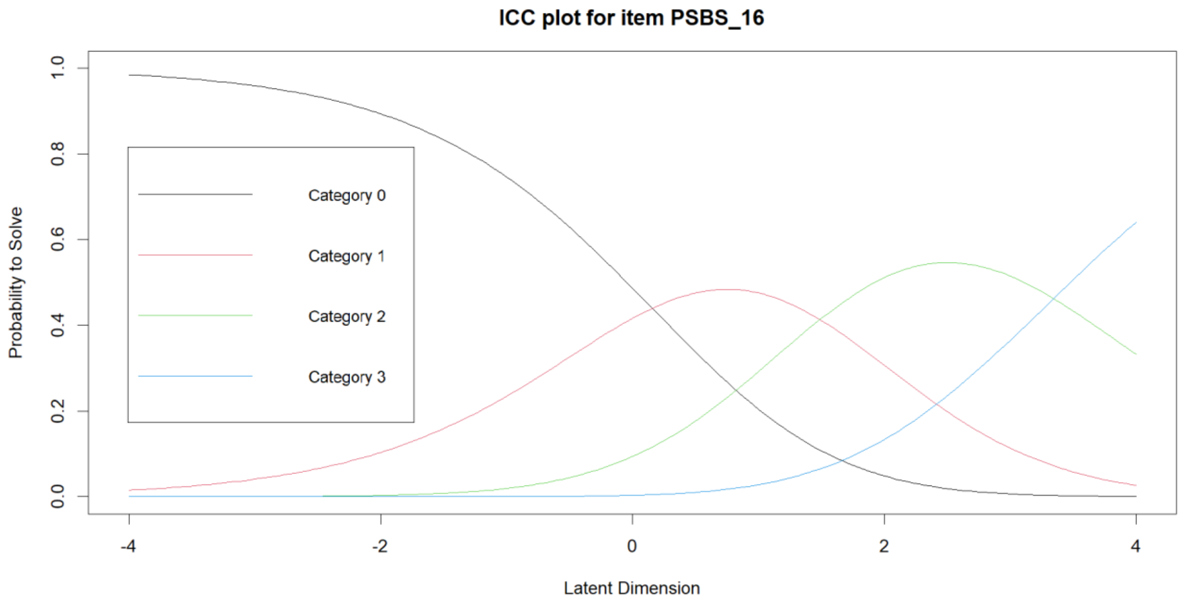
Appendix K – Item characteristic curves (ICC) for all items of the Paranormal and Supernatural Beliefs Scale (PSBS; Dean et al., 2021)

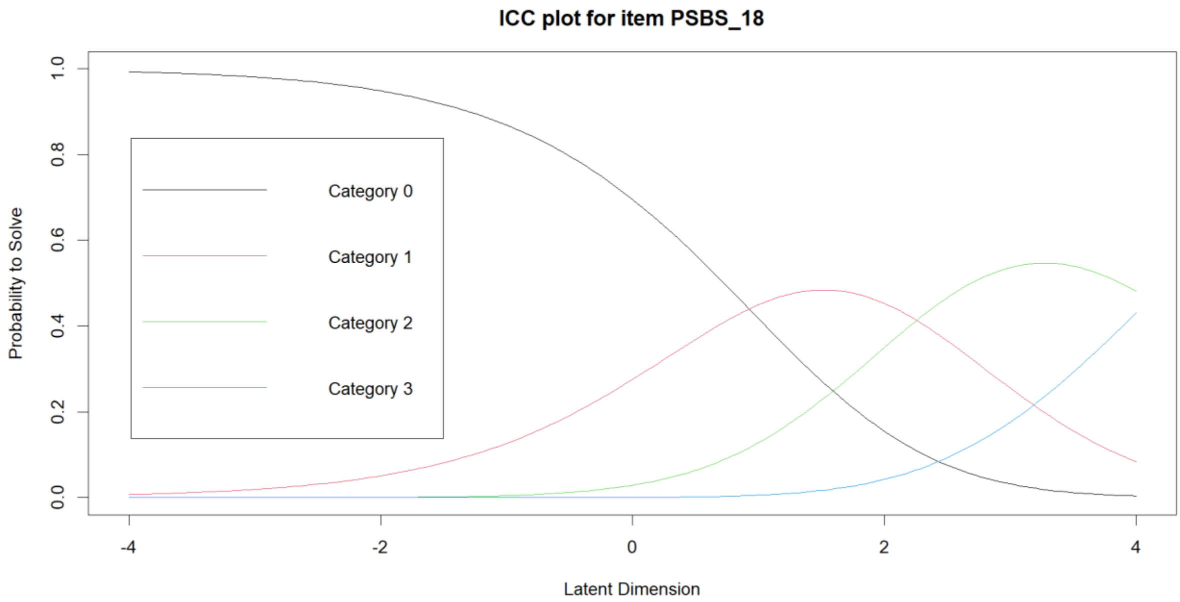


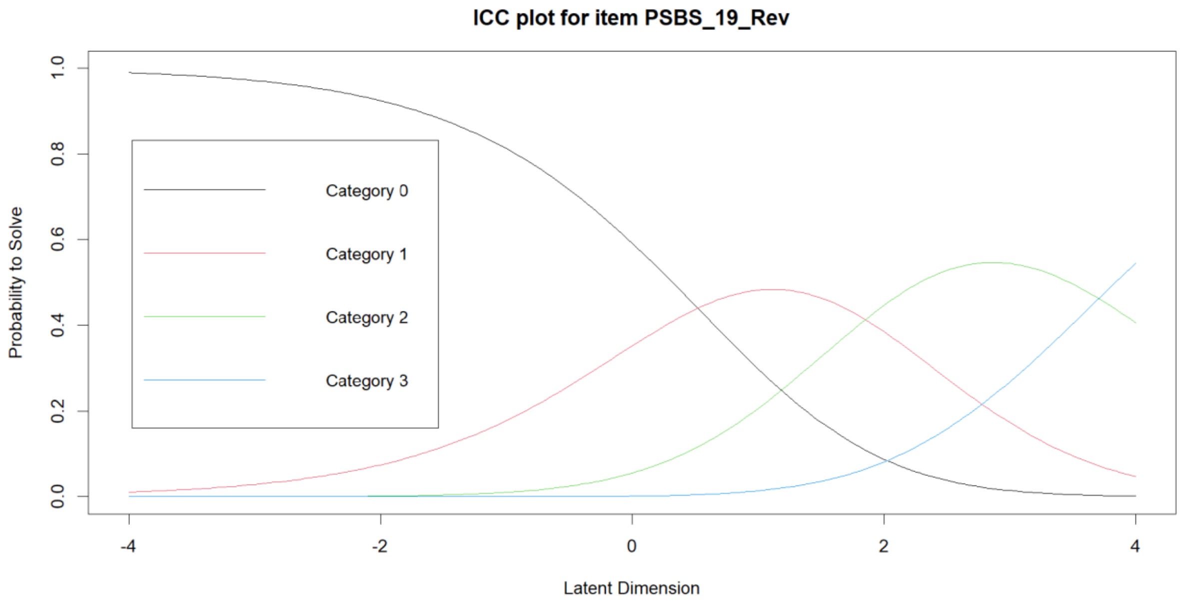


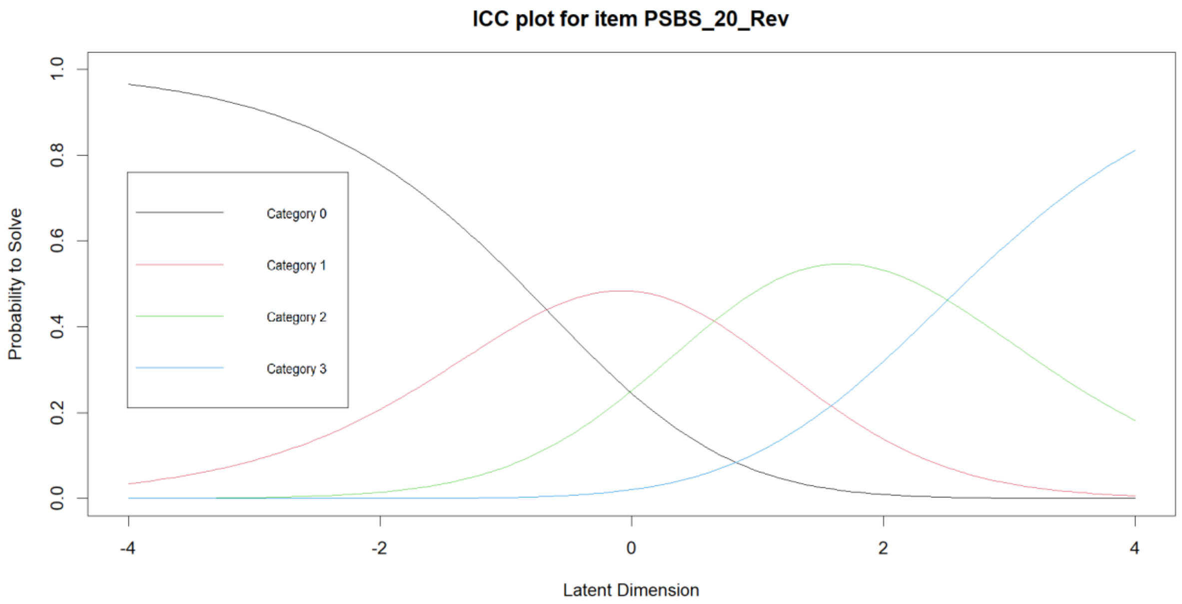

ICC plot for item PSBS_9
Probability to Solve
Latent Dimension
Category 0
Category 1
Category 2
Category 3
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4

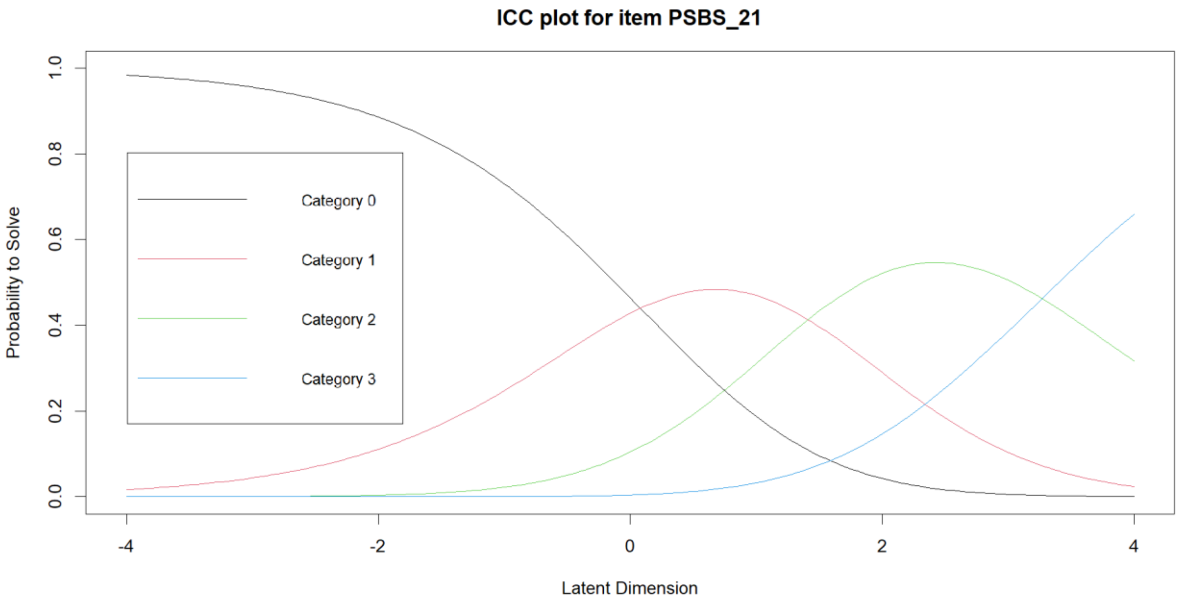

ICC plot for item PSBS_11
Probability to Solve
Latent Dimension
Category 0
Category 1
Category 2
Category 3
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4

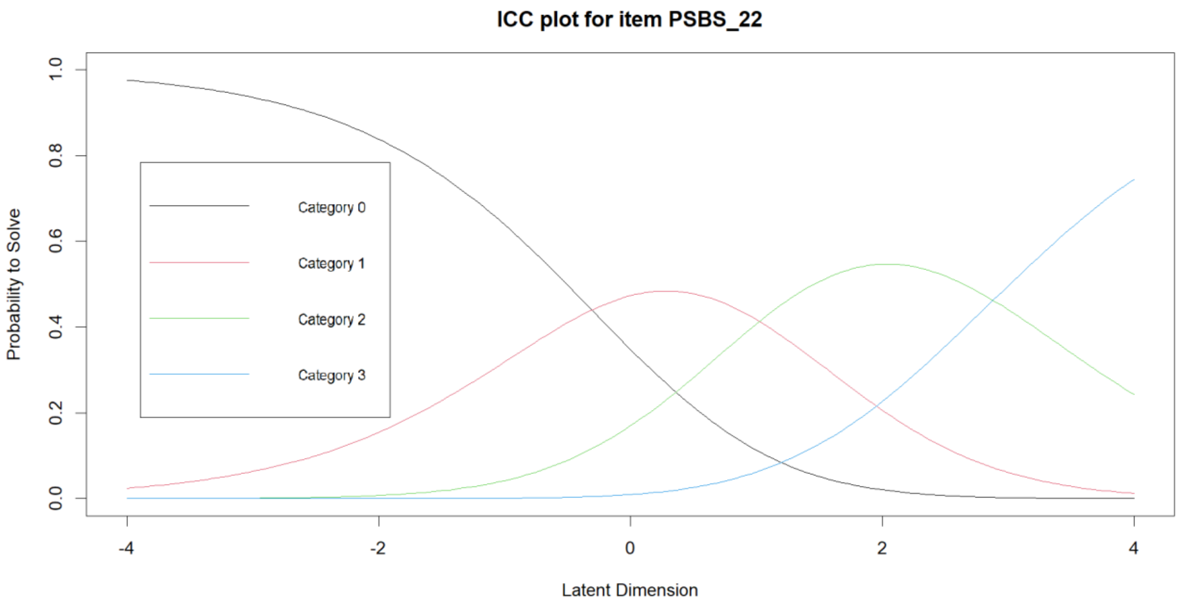

ICC plot for item PSBS_16
Probability to Solve
Latent Dimension
Category 0
Category 1
Category 2
Category 3
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4

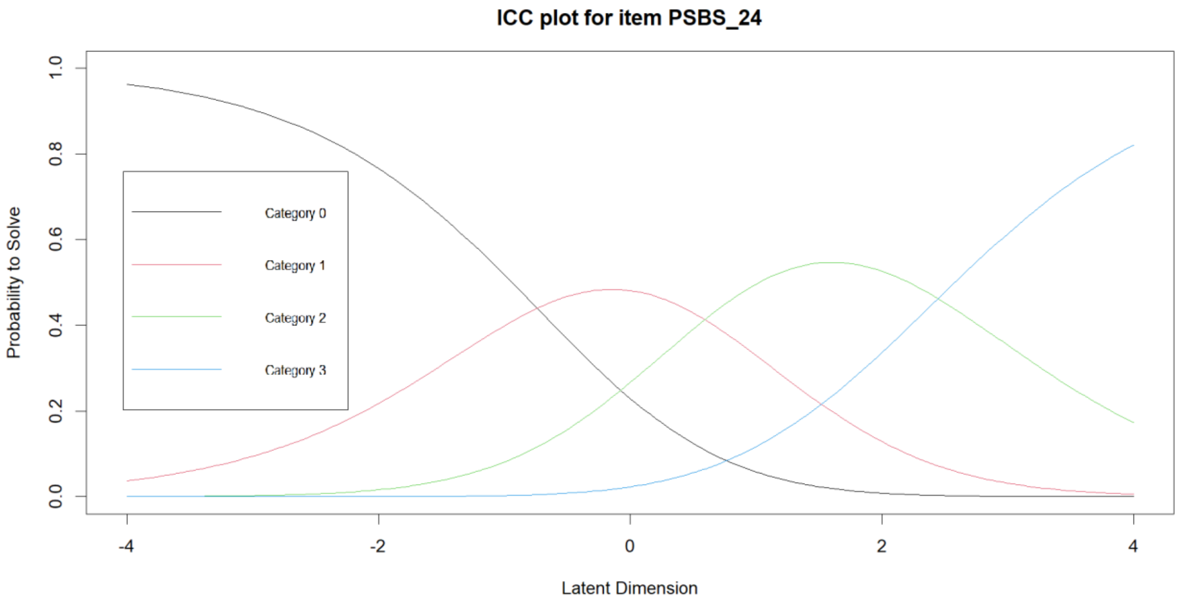

ICC plot for item PSBS_18
Probability to Solve
Latent Dimension
Category 0
Category 1
Category 2
Category 3
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4

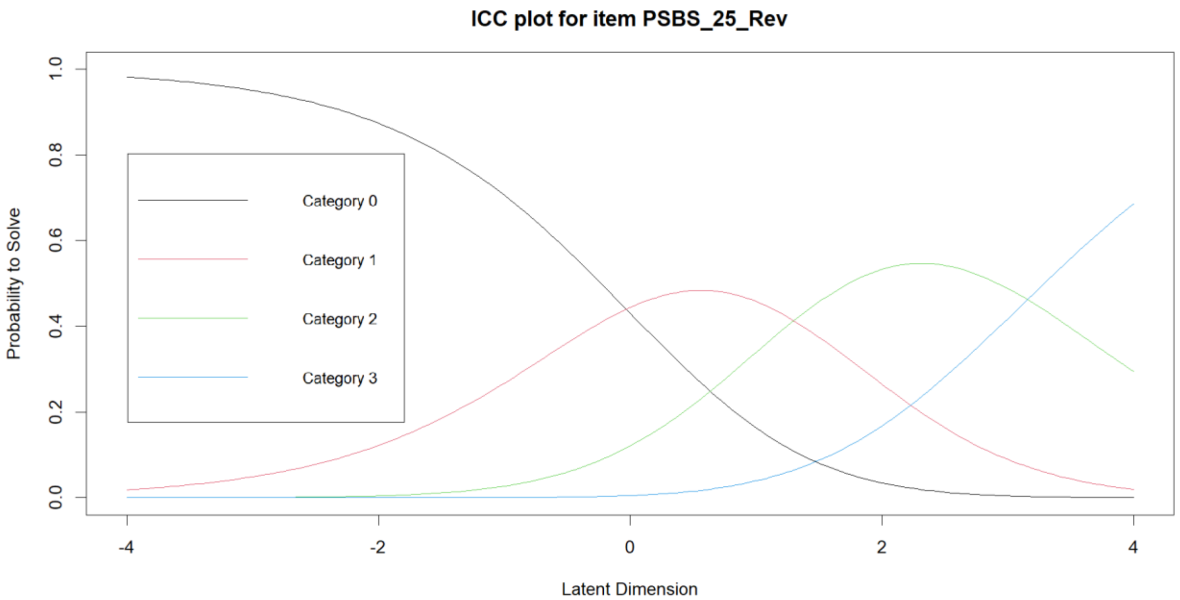

ICC plot for item PSBS_19_Rev
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension
Category 0
Category 1
Category 2
Category 3

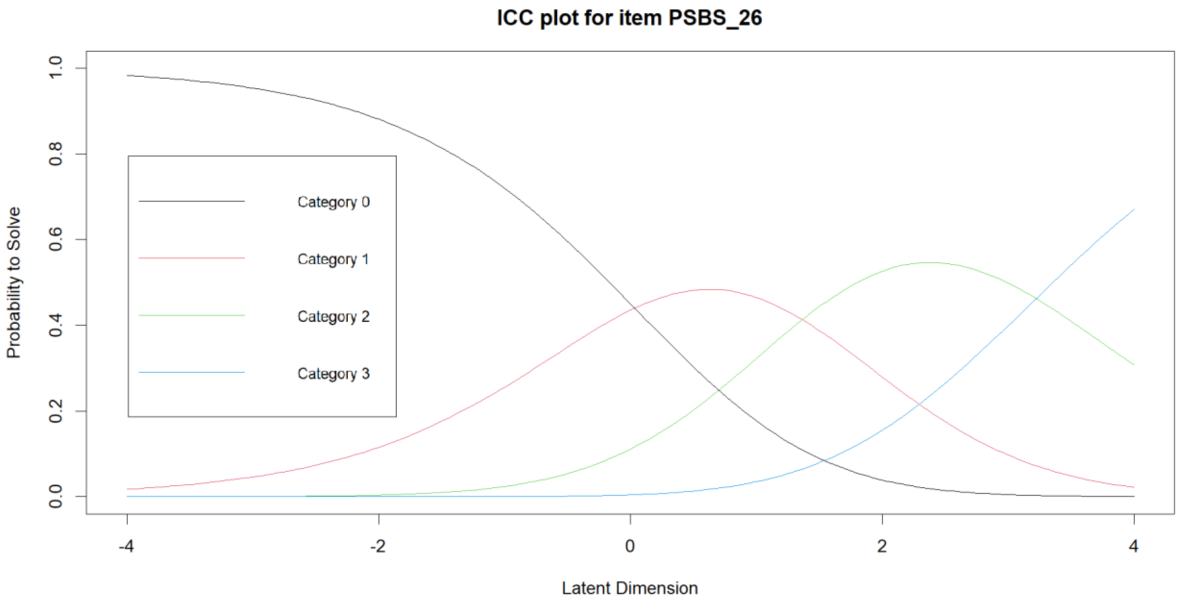

ICC plot for item PSBS_20_Rev
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
Latent Dimension
Category 0
Category 1
Category 2
Category 3


ICC plot for item PSBS_21
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
Category 0
Category 1
Category 2
Category 3
-4
-2
0
2
4
Latent Dimension


ICC plot for item PSBS_22
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
Category 0
Category 1
Category 2
Category 3
-4
-2
0
2
4
Latent Dimension


ICC plot for item PSBS_24
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
Category 0
Category 1
Category 2
Category 3
-4
-2
0
2
4
Latent Dimension


ICC plot for item PSBS_25_Rev
Probability to Solve
0.0
0.2
0.4
0.6
0.8
1.0
Category 0
Category 1
Category 2
Category 3
-4
-2
0
2
4
Latent Dimension


ICC plot for item PSBS_26
Probability to Solve
1.0
0.8
0.6
0.4
0.2
0.0
Category 0
Category 1
Category 2
Category 3
-4
-2
0
2
4
Latent Dimension

Appendix L – Dean et al.'s (2021) Paranormal and Supernatural Belief Scale

Paranormal and Supernatural Belief Scale

The following questions concern your beliefs about paranormal phenomena. There are no right or wrong answers. This is a sample of your own beliefs and attitudes. Please be honest in your responses. Thank you.

0 = Strongly Disagree 1 = Disagree 2 = Agree 3 = Strongly Agree

1. Your mind or soul can leave your body.
2. If you break a mirror, you will have bad luck.
3. It is possible to be reincarnated.
4. Mind reading is possible.
5. A person's star sign can have a direct influence on their personality.
6. Fairies and similar beings are real.

7*. Fortune tellers' predictions are typically based on guesswork.

8*. Reports of an apparent sixth sense are generally based on fantasies.

9. Some health conditions can be treated with psychic healing.
10. In some cultures, shamans or "witch doctors" exercise powers we cannot explain.
11. Having a dream that comes true is not just a coincidence.

12*. Communicating with spirits or other supernatural entities through a Ouija board is not possible.

13. It is possible to become possessed by an evil supernatural entity.

Note. Items 7, 8, and 12 are reverse scored.

Appendix M – Peters et al. (2004) Delusions Inventory

P.D.I.-21

This questionnaire is designed to measure beliefs and vivid mental experiences. We believe that they are much more common than has previously been supposed, and that most people have had some such experiences during their lives. Please answer the following questions as honestly as you can. There are no right or wrong answers, and there are no trick questions.

Please note that we are NOT interested in experiences people may have had when under the influence of drugs.

IT IS IMPORTANT THAT YOU ANSWER *ALL* QUESTIONS.

For the questions you answer YES to, we are interested in:
(a) how distressing these beliefs or experiences are
(b) how often you think about them; and
(c) how true you believe them to be.
On the right hand side of the page we would like you to circle the number which corresponds most closely to how distressing this belief is, how often you think about it, and how much you believe that it is true.
If you answer NO please move on to the next question.

Example

Do you ever feel as if people are reading your mind ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

Do you ever feel as if you could read other people's minds ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

1) Do you ever feel as if people seem to drop hints about you or say things with a double meaning ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

2) Do you ever feel as if things in magazines or on TV were written especially for you ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

3) Do you ever feel as if some people are not what they seem to be ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

4) Do you ever feel as if you are being persecuted in some way ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

5) Do you ever feel as if there is a conspiracy against you ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

6) Do you ever feel as if you are, or destined to be someone very important ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

7) Do you ever feel that you are a very special or unusual person ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

8) Do you ever feel that you are especially close to God ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

9) Do you ever think people can communicate telepathically ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

10) Do you ever feel as if electrical devices such as computers can influence the way you think ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

11) Do you ever feel as if you have been chosen by God in some way ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

12) Do you believe in the power of witchcraft, voodoo or the occult ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

13) Are you often worried that your partner may be unfaithful ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

14) Do you ever feel that you have sinned more than the average person ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

15) Do you ever feel that people look at you oddly because of your appearance ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

16) Do you ever feel as if you had no thoughts in your head at all ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

17) Do you ever feel as if the world is about to end ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

18) Do your thoughts ever feel alien to you in some way ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

19) Have your thoughts ever been so vivid that you were worried other people would hear them ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

20) Do you ever feel as if your own thoughts were being echoed back to you ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

21) Do you ever feel as if you are a robot or zombie without a will of your own ?

NO YES

(please circle)

Not at all distressing				Very distressing
1	2	3	4	5
Hardly ever think about it				Think about it all the time
1	2	3	4	5
Don't believe it's true				Believe it is absolutely true
1	2	3	4	5

Appendix N – Wilson et al.'s (1996) Dysexecutive Questionnaire

The following questions look at some of the difficulties that people sometimes experience.

Please read the following statements, and rate them on the five-point scale according to your personal experiences.

1. I have problems understanding what other people mean unless they keep things simple and straightforward
2. I act without thinking, doing the first thing that comes to mind
3. I sometimes talk about events or details that never actually happened, but I believe did happen
4. I have difficulty thinking ahead or planning for the future
5. I sometimes get over-excited about things and can be a bit 'over-the-top' at these times
6. I get events mixed up with each other, and get confused about the correct order of events
7. I have difficulty realizing the extent of my problems and am unrealistic about the future
8. I seem lethargic, or unenthusiastic about things
9. I do or say embarrassing things when in the company of others
10. I really want to do something one minute, but couldn't care less about it the next
11. I have difficulty showing emotion
12. I lose my temper at the slightest thing
13. I seem unconcerned about how I should behave in certain situations
14. I find it hard to stop repeating saying or doing things once started
15. I tend to be very restless and 'can't sit still' for any length of time
16. I find it difficult to stop doing something even if I know I shouldn't
17. I will say one thing, but do something different
18. I find it difficult to keep my mind on something, and am easily distracted
19. I have trouble making decisions, or deciding what I want to do
20. I am unaware of, or unconcerned about, how others feel about my behaviour

0= never 1= occasionally 2= sometimes 3= fairly often 4= very often

Appendix O – Raine & Benishay's (1995) Schizotypal Personality Questionnaire – Brief

The following statements and questions relate to certain life experiences and beliefs that some people may have. There are no right or wrong answers, and there are no trick questions. Please be honest in your responses.

Thank you.

0 = Strongly Disagree 1 = Disagree 2 = Neutral 3 = Agree 4 = Strongly Agree

1. I feel I have to be on my guard even with friends.
2. Have you found that it is best not to let other people know too much about you?
3. Do you feel that you are unable to get "close" to people?
4. Do you often have to keep an eye out to stop people from taking advantage of you?
5. I tend to keep my feelings to myself.
6. Do you often pick up hidden threats or put-downs from what people say or do?
7. Some people find me a bit vague and elusive during a conversation.
8. Some people think that I am a very bizarre person.
9. I am an odd, unusual person.
10. People sometimes comment on my unusual mannerisms and habits.
11. I sometimes use words in unusual ways.
12. People sometimes find me aloof and distant.
13. I feel very uneasy talking to people I do not know well.
14. I feel very uncomfortable in social situations involving unfamiliar people.
15. I tend to keep in the background on social occasions.
16. I find it hard to communicate clearly what I want to say to people.
17. Have you ever noticed a common event or object that seemed to be a special sign for you?
18. Are you sometimes sure that other people can tell what you are thinking?
19. Have you ever had the sense that some person or force is around you, even though you cannot see anyone?
20. Do you ever suddenly feel distracted by distant sounds that you are not normally aware of?
21. When shopping do you get the feeling that other people are taking notice of you?
22. Have you had experiences with astrology, seeing the future, UFOs, ESP, or a sixth sense?

Appendix P – Dennis & Vander Wal's (2010) Cognitive Flexibility Inventory

The following statements look at some of the experiences people may have when making decisions in difficult situations.

Please read the statements carefully and rate them on the seven-point scale according to your personal experience.

Thank you.

1 = Strongly Disagree 2 = Disagree 3 = Somewhat Disagree 4 = Uncertain 5 = Somewhat Agree 6 = Agree 7 = Strongly Agree

1. I am good at "sizing up" situations.
2. I have a hard time making decisions when faced with difficult situations.
3. I consider multiple options before making a decision.
4. When I encounter difficult situations, I feel like I am losing control.
5. I like to look at difficult situations from many different angles.
6. I seek additional information not immediately available before attributing causes to behaviour.
7. When encountering difficult situations, I become so stressed that I cannot think of a way to resolve the situation.
8. I try to think about things from another person's point of view.
9. I find it troublesome that there are so many different ways to deal with difficult situations.
10. I am good at putting myself in others' shoes.
11. When I encounter difficult situations, I just don't know what to do.
12. It is important to look at difficult situations from many angles.
13. When in difficult situations, I consider multiple options before deciding how to behave.
14. I often look at a situation from different view-points.
15. I am capable of overcoming the difficulties in life that I face.
16. I consider all the available facts and information when attributing causes to behaviour.
17. I feel I have no power to change things in difficult situations.
18. When I encounter difficult situations, I stop and try to think of several ways to resolve it.
19. I can think of more than one way to resolve a difficult situation I'm confronted with.
20. I consider multiple options before responding to difficult situations.

Note. Items 2, 4, 7, 9, 11 & 17 are reverse scored.

Appendix Q – Buhr & Dugas' (2002) English translation of the Intolerance of Uncertainty Scale (Freeston et al., 1994)

The following statements describe how people may react to the uncertainties of life. Please use the scale provided to describe to what extent each item is characteristic of you.

There are no right or wrong answers, please be honest in your responses.

Thank you.

1 = not at all characteristic of me 2 = slightly characteristic of me 3 = moderately characteristic of me 4 = very characteristic of me 5 = extremely characteristic of me

1	Uncertainty stops me from having a strong opinion.
2	Being uncertain means that a person is disorganized.
3	Uncertainty makes life intolerable.
4	It's unfair having no guarantees in life.
5	My mind can't be relaxed if I don't know what will happen tomorrow.
6	Uncertainty makes me uneasy, anxious, or stressed.
7	Unforeseen events upset me greatly.
8	It frustrates me not having all the information I need.
9	Uncertainty keeps me from living a full life.
10	One should always look ahead so as to avoid surprises.
11	A small unforeseen event can spoil everything, even with the best planning.
12	When it's time to act, uncertainty paralyses me.
13	Being uncertain means that I am not first rate.
14	When I am uncertain, I can't go forward.
15	When I am uncertain, I can't function very well.
16	Unlike me, others seem to know where they are going with their lives.
17	Uncertainty makes me vulnerable, unhappy, or sad.
18	I always want to know what the future has in store for me.
19	I can't stand being taken by surprise.
20	The smallest doubt can stop me from acting.
21	I should be able to organize everything in advance.
22	Being uncertain means that I lack confidence.
23	I think it's unfair that other people seem to be sure about their future.
24	Uncertainty keeps me from sleeping soundly.
25	I must get away from all uncertain situations.
26	The ambiguities in life stress me.
27	I can't stand being undecided about my future.

Appendix R – Roberts et al.'s (2011) Detail and Flexibility Questionnaire

The following statements relate to your beliefs and feelings about your own behaviour. Read each statement carefully and use the scale provided to indicate how much you agree or disagree with each statement. There are no right or wrong answers, and no trick questions. Please be honest in your responses.

Thank you.

1 = Strongly Disagree 2 = Disagree 3 = Slightly Disagree 4 = Slightly Agree 5 = Agree 6 = Strongly Agree

1. I get angry if people do not do things my way
2. I sometimes bore others as I go on to an excess about somethings
3. I get upset if other people disturb my plans for the day by being late
4. I have difficulty making decisions
5. When others suggest a new way of doing things, I get upset or unsettled
6. I find it difficult to remember the story line in films, plays or books, but can remember specific scenes in great detail
7. Once I get into an emotional state, eganger or sadness, it is very difficult to soothe myself
8. I spend as much time on more or less important tasks
9. I like to make plans about complex arrangements, eg journeys and work projects
10. I can get hung up on details when reading rather than understanding the gist
11. I have high levels of anxiety/discomfort: I can see/feel/taste that things might not be quite right
12. I tend to focus on one thing at a time and get it out of proportion to the total situation
13. I like doing things in a particular order or routine
14. I can get lost in details and forget the real purpose of a task
15. I can be called stubborn or single minded as it is difficult to shift from one point of view to another
16. I find it difficult to do several things at once (multitasking)
17. I need clarity and rules when facing a new situation. Without rules, I easily feel lost
18. I find it hard to see different perspectives of a situation
19. I get very distressed if plans get changed at the last minute
20. I can get overwhelmed by too many details
21. I dislike change
22. I depend on others to help me get things into perspective, as I tend to have a rather blinkered view on things in my life
23. I often feel vulnerable and unsafe as I am unable to see threats (or opportunities) that are out of my field of vision
24. I find it hard to write concisely: I often overrun word limits and find it difficult to decide which details can be left out

Notes:
Cognitive rigidity subscale — odd numbered items.
Attention to detail subscale — even numbered items.