Paranormal psychic believers and skeptics: a large-scale test of the cognitive differences hypothesis

Stephen J. Gray¹ · David A. Gallo¹

Published online: 26 October 2015
© Psychonomic Society, Inc. 2015

Abstract Belief in paranormal psychic phenomena is widespread in the United States, with over a third of the population believing in extrasensory perception (ESP). Why do some people believe, while others are skeptical? According to the cognitive differences hypothesis, individual differences in the way people process information about the world can contribute to the creation of psychic beliefs, such as differences in memory accuracy (e.g., selectively remembering a fortune teller's correct predictions) or analytical thinking (e.g., relying on intuition rather than scrutinizing evidence). While this hypothesis is prevalent in the literature, few have attempted to empirically test it. Here, we provided the most comprehensive test of the cognitive differences hypothesis to date. In 3 studies, we used online screening to recruit groups of strong believers and strong skeptics, matched on key demographics (age, sex, and years of education). These groups were then tested in laboratory and online settings using multiple cognitive tasks and other measures. Our cognitive testing showed that there were no consistent group differences on tasks of episodic memory distortion, autobiographical memory distortion, or working memory capacity, but skeptics consistently outperformed believers on several tasks tapping analytical or logical thinking as well as vocabulary. These findings demonstrate cognitive similarities and differences between these groups and suggest that differences in analytical thinking and conceptual knowledge might contribute to the development of psychic beliefs. We also found that psychic belief was associated with greater life satisfaction, demonstrating

benefits associated with psychic beliefs and highlighting the role of both cognitive and noncognitive factors in understanding these individual differences.

Keywords False memory · Individual differences · Memory · Working memory · Problem solving

Many people in the United States believe in the existence of psychic phenomena, such as extrasensory perception (42 %), telepathy (31 %), or clairvoyance (26 %; Gallup, 2005), while many others reject such phenomena as illogical or nonscientific. Understanding why some people strongly believe in psychic phenomena whereas others are strongly skeptical may provide insights into the factors that drive individual differences in scientific thinking and reasoning more generally. Research suggests that many factors are involved in the propensity for various kinds of paranormal beliefs, including sociocultural traditions (e.g., being raised by a caregiver with similar beliefs; Braswell, Rosengren, & Berenbaum, 2012) and the psychological benefits that such beliefs might afford (e.g., enhanced meaning in life; see Kennedy, Kanthamani, & Palmer, 1994; Parra & Corbetta, 2013). These findings suggest that paranormal beliefs are influenced by a variety of individual differences, and they raise the question as to whether other factors – such as individual differences in cognition – also contribute to the propensity for paranormal beliefs. Cognitive differences may be particularly relevant to beliefs in psychic phenomena (e.g., ESP or clairvoyance), because unlike other kinds of supernatural beliefs (e.g., ghosts or UFOs), psychic beliefs have a metacognitive component: By definition, they require thinking about the cognitive abilities and limitations of the human mind.

With the current research we investigated the cognitive differences hypothesis that has been put forth in the literature,

✉ Stephen J. Gray
sjgray@uchicago.edu

¹ Department of Psychology, University of Chicago,
Chicago, IL 60637, USA

or the idea that individual differences in psychic beliefs are related to differences in the way people tend to process information about the world (e.g., Blackmore, 1992; see Irwin, 2009). It is important to note up front that such cognitive differences – if they exist – need not be considered inherently “good” or “bad,” nor would they necessarily imply differences in overall cognitive ability or potential for success. Indeed, prior studies have failed to find consistent links between various kinds of paranormal beliefs (including psychic beliefs) and global measures of intelligence (e.g., Smith, Foster, & Stovin, 1998; Watt & Wiseman, 2002; Musch & Ehrenberg, 2002) or academic achievement (e.g., Messer & Griggs, 1989; Tobacyk, Miller, & Jones, 1984), providing little evidence for differences in overall cognitive ability. Rather than making such global characterizations of believers and skeptics, our goal with this research was to determine the extent that individual differences in paranormal beliefs are related to differences in information processing within specific cognitive domains. Such differences might be driven by domain-specific abilities, or they might be driven by more flexible information processing styles that, in turn, could be influenced by other factors (e.g., speed–accuracy trade-offs, personality characteristics, motivation). Regardless of the cause, such cognitive differences might contribute to the development and reinforcement of psychic beliefs, but few studies have tested for cognitive differences between those who believe in psychic phenomena and those who do not.

With respect to specific cognitive domains, our primary focus was episodic memory accuracy, or one’s susceptibility to memory errors and distortions (for reviews of individual differences in various memory distortions, see Gallo, 2010; Zhu et al., 2010). According to the *memory distortion hypothesis*, an enhanced susceptibility to memory errors and biases increases the likelihood that some people will selectively recall facts or events that support their psychic beliefs (see also Blackmore, 1992). For example, a believer may be biased to selectively remember a fortuneteller’s correct predictions or to distort their own memory to conform to these predictions, helping to create and reinforce their beliefs in such psychic phenomena. A greater propensity for memory distortions also might increase the likelihood of believers misattributing normal experiences to psychic ones. One might dream about a prior experience and then later misremember the dream as preceding the experience (e.g., a psychic vision). Similarly, one might have an initial conversation with a friend about their feelings, and then later forget this conversation and instead misattribute a deep understanding of their friend’s feelings to a paranormal sense of intuition or ESP. In general, according to this hypothesis, an increased susceptibility to memory distortions could affect one’s sense of reality in ways that could support an individual’s psychic beliefs.

To our knowledge, no prior research has investigated the possible link between memory distortion and psychic beliefs

in particular, but an influential pair of studies by McNally and colleagues is relevant to the memory distortion hypothesis. These studies used the DRM false memory illusion (Roediger & McDermott, 1995), which measures people’s propensity to falsely remember nonpresented words after attempting to memorize lists of semantically associated words. Clancy, McNally, Schacter, Lenzenweger, and Pitman (2002) found that individuals who claimed to have memories of being abducted by extraterrestrial aliens were more susceptible to the DRM memory illusion compared to control participants (i.e., people who did not report being abducted by aliens). Similarly, Meyersburg, Bogdan, Gallo, and McNally (2009) found that individuals who claimed to have memories from past lives (e.g., reincarnation) also were more prone to the DRM illusion than control participants. These studies demonstrate a link between false memories and paranormal beliefs, suggesting that psychic beliefs also might be related to memory distortion. However, a limitation to this conclusion is that participants in these studies were intentionally recruited because they claimed to have memories of paranormal experiences. Assuming that these paranormal experiences did not actually happen, these participant recruitment methods conflated the two factors of interest here (i.e., propensity for memory distortion and paranormal beliefs). In order to test the hypothesis that paranormal belief is related to memory distortion more generally, one would need to recruit participants solely on the basis of paranormal belief and then test for differences in memory distortion.

Although no studies have investigated the potential link between memory distortion and psychic beliefs, a few studies have investigated the link between more general paranormal beliefs and memory accuracy, and the results have been mixed. Blackmore and Rose (1997) and Rose and Blackmore (2001) found no relationship between paranormal beliefs and memory accuracy on a task in which college undergraduates had to differentiate between seen and imagined objects in memory. French, Santomauro, Hamilton, Fox, and Thalbourne (2008) did not find a significant relationship between personally reported extraterrestrial alien contact experiences and false memory on the DRM task (in contrast to Clancy et al., 2002) but did find a weak positive relationship between false memory and self-proclaimed paranormal experiences and abilities. Finally, Corlett and colleagues (2009) found that magical thinking (including beliefs in paranormal phenomena) was positively related to the DRM illusion in a sample of college undergraduates. It is unclear why these relationships were only sometimes found, but the use of relatively small sample sizes and a limited number of memory tasks make this literature difficult to interpret.

Despite the unclear pattern of results from these studies, there have been consistent correlations between beliefs in various paranormal phenomena and other personality characteristics that, in turn, have been linked to laboratory measures of

memory distortion. For example, Glicksohn and Barrett (2003) found a relationship between paranormal beliefs, absorption, and hallucinatory experiences, and Wilson and French (2006) linked paranormal beliefs to dissociativity, absorption, fantasy proneness, and hypnotic suggestibility. In general, individual differences in absorption and dissociative experiences have been found to correlate with propensity for memory distortion (see Eisen & Lynn, 2001; Gallo, 2010). If these metrics are correlated with both false memories and paranormal beliefs, then perhaps there also is a direct relationship between memory distortion and psychic beliefs.

In addition to memory distortion, the current research also investigated the potential link between individual differences in psychic beliefs and analytical or critical thinking. According to the *analytical thinking hypothesis*, differences in analytical or critical thinking could promote psychic beliefs, to the extent that believers are less likely than skeptics to scrutinize evidence that may not support their beliefs. Consistent with this hypothesis, several studies have found a negative relationship between psychic beliefs and syllogistic reasoning tasks thought to tap analytical thinking (e.g., Roberts & Seager, 1999; Watt & Wiseman, 2002; Wiseman & Watt, 2006). In general, this research on psychic beliefs is consistent with studies linking other kinds of paranormal beliefs to less analytic or critical thinking (see Gervais & Norenzayan, 2012; Pennycook, Cheryne, Sleli, Koehler, & Fugelsang, 2012), including less critical evaluation of hypothetical arguments (Stanovich & West, 1998; Svedholm & Lindeman, 2012) or an increased likelihood of endorsing conspiracy theories that most people reject based on careful scrutiny of the evidence (Bruder, Haffke, Neave, Nouripanah, & Imhoff, 2013; Lobato, Mendoza, Sims, & Chin, 2014). Although group differences in analytical thinking and logic tasks have not always been observed (compare Dagnall, Drinkwater, Parker, & Rowley, 2014, to Rogers, Davis, & Fisk, 2009), these represent some of the most reliable cognitive difference between paranormal believers and skeptics observed to date.

Current study

The primary goal of the current study was to provide a comprehensive test of the memory distortion hypothesis, testing the prediction that individual differences in memory accuracy and distortion will be associated with paranormal psychic beliefs. We tested this hypothesis using a combination of laboratory and online tasks as well as multiple memory measures that included both episodic memory and autobiographical memory tasks. We also tested working memory, which has been linked to individual differences in memory distortion in other studies (Watson, Bunting, Poole, & Conway, 2005; Unsworth & Brewer, 2009). A secondary goal of the current

research was to investigate the potential links between psychic beliefs and measures of analytical thinking and personality characteristics. Given that prior research has linked these measures to paranormal beliefs in general, we set out to replicate these findings and extend them to psychic beliefs in particular.

In all three of our studies, we measured individual differences in psychic beliefs in large online samples using a self-report questionnaire coupled with an open-ended question about one's reasons for their belief or skepticism. Independent online samples were recruited for each study, from which groups of strong believers and skeptics were recruited for follow-up cognitive testing. For each of our studies, groups were matched on key demographics (age, sex, years of education), and in two of our studies we also matched believers and skeptics on academic achievement (self-reported GPA). These matching procedures ensured that any observed cognitive differences could be attributed to specific cognitive domains as opposed to more general intellectual functioning or academic achievement.

We took a between-groups approach, rather than treating paranormal belief as a continuous individual difference variable, for both theoretical and practical reasons. On the theoretical side, while beliefs in psychic phenomena can range in confidence or strength, it could be argued that there is a categorical difference between strongly believing that paranormal phenomena exist and strongly believing that they do not exist. Similar to other research in this field (e.g., Riekk, Lindeman, Aleneff, Halme, & Nuortimo, 2013), our research question was aimed at comparing strong believers and strong disbelievers while excluding those that do not strongly believe but would remain open-minded or agnostic. To the extent that believers and skeptics differ on cognitive measures, an extreme groups approach should be able to detect this difference. On the practical side, our approach involved multiple sessions of rigorous cognitive testing that would be unfeasible in a larger sample that included more agnostic individuals. Instead, we used a multistaged procedure to recruit targeted groups of strong believers and strong skeptics, thereby increasing the likelihood of adequate statistical power to detect group differences on our cognitive measures. This between-groups approach represents an efficient way to test for possible differences between believers and skeptics across multiple cognitive domains, even though this approach obviously oversimplifies the complexity of individual beliefs.

We conducted three independent studies so that we could administer different kinds of cognitive tasks and also attempt to replicate key findings across independent samples. Memory accuracy and distortion was measured with three tasks. (1) A version of the Deese-Roediger-McDermott task (DRM; Roediger & McDermott, 1995), which tested participant's memory for studied words as well as false memory for nonstudied but associated words. (2) A criterial recollection task (CRT), which required participants to accurately recollect

different kinds of information (e.g., font color or pictures) associated with studied words (Gallo, 2013). (3) An imagination inflation task (IIT), which used guided imagery to bias estimates of the occurrence of childhood events from autobiographical memory (Garry, Manning, Loftus, & Sherman, 1996). We assessed working memory with the reading-span (RSPAN) and operation-span (OSPAN) tasks (Daneman & Carpenter, 1980; Turner & Engle, 1989).

In addition to these memory tasks, we included several measures tapping different aspects of analytical or critical thinking across the three studies. (1) The *Shipley Institute of Living Scale* (Zachary, 1986), which contains a logical reasoning component in which participants are required to look at a pattern of letters, numbers, or words and indicate the next item in the pattern. The Shipley also includes a vocabulary test. (2) An argument evaluation task (AET; Stanovich & West, 1997), which requires participants to evaluate the quality of the positions and arguments that other people make when debating various public issues. (3) The remote associations test (Mednick, 1962), in which participants are given three words (e.g. *stalk*, *trainer*, *king*) and required to solve the puzzle by identifying the fourth word that connects all three (*lion*). (4) A conspiracy theories questionnaire, based on items from Oliver and Wood (2014), which assessed the likelihood that participants would endorse conspiracy theory statements about current and historical events relative to matched control items as a baseline.

Finally, we included several personality and self-report measures that have previously been linked to propensity for memory distortion. These included the Dissociative Experiences Questionnaire–Comparative (DES-C; Wright & Loftus, 1999), the Tellegen Absorption Scale (TAS; Tellegen & Atkinson, 1974), and the Need for Cognition Questionnaire (Cacioppo, Petty, & Kao, 1984). We also asked participants a few exploratory questions to target other aspects of their worldview. One set of questions assessed their beliefs in Darwin’s theory of biological evolution. This set of questions was included to determine whether beliefs in paranormal psychic phenomena generalized to a rejection of all kinds of scientific beliefs, or instead, whether they were specific to the existence of paranormal phenomena. Another question asked about overall life satisfaction, as research has found that paranormal beliefs are associated with greater meaning in life and well-being (Kennedy et al., 1994; Parra & Coretta, 2013), suggesting that these kinds of beliefs might have beneficial properties in some individuals.

Method

In this study, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012)

Participant recruitment

Each of the three studies had the same overall structure. In the initial screening stage, hundreds of participants took part in an online procedure that measured psychic beliefs with a modified version of the Australian Sheep-Goat Scale (ASGS; Thalbourne & Delin, 1993), which is described more thoroughly below. Individuals with the highest scores (strongest believers) and lowest scores (strongest skeptics) were invited to participate in additional cognitive testing, provided that they were between the ages of 18 and 35 and had not experienced severe brain injury/disease or other form of cognitive decline (self-report). In all three studies we aimed to match these groups of skeptics and believers on age, sex, and years of education, and we further matched for self-reported GPA in Studies 2 and 3. In Study 1, the initial screening phase was advertised on Craigslist in the Chicago area, and participants who passed the screening stage were recruited to come into the laboratory and complete the cognitive tasks and other measures described below. In Studies 2 and 3, the initial screening phase was advertised nationally on Amazon Mechanical Turk, and participants who passed the screening stage were recruited to take the cognitive tasks and other measures online. In these studies, we sought to replicate and extend the findings from Study 1 in a broader online population. (See Fig. 1 for a graphical representation of the study procedures.)

In each study we tested at least 40 believers and 40 skeptics for the final group comparisons. A power analysis of Meyersburg et al. (2009), who found differences between believers in past lives and nonbelievers in the DRM task (i.e., one of the primary measures in our study), revealed that a sample size of 30 should be sufficient to detect a similar group difference between psychic believers and skeptics (80 % power to detect a mean group difference of 14 % in false recall, $SD = .22$, $\alpha = .05$, one-tailed). Our final sample sizes in each study were larger than this, in part because we recognized that a more common between-groups comparisons in the DRM literature (i.e., younger vs. older adults) typically uses more participants than had been used in Meyersburg et al. (2009). Given that 40 participants per group would be a relatively large sample size compared to the expansive aging literature with this task, we reasoned that testing this number of participants (or more) in each group of the current study would give us sufficient power to detect an effect at least as large as the aging effect on memory distortion, thereby providing a concrete anchor for evaluating our current effects. We also overrecruited participants in our online sampling procedure to ensure a sufficient number would take part in the follow-up sessions. Sampling also was constrained by our group matching procedures for each study, as described below.

Study 1

Study 2

Study 3

Fig. 1 Note. A graphical representation of study procedures. CRT = criterial recollection task; TAS = *Tellegen Absorption Scale*; DES-C = dissociative experiences scale-comparative; AET = argument evaluation task; IIT =

imagination inflation task; RSPAN/OSPAN = reading span/operation span working memory tasks; RAT = remote associations task

Study 1 sample

In this study, 731 (256 male, mean age = 27.8 years before exclusion) individuals were initially screened and recruited via Chicago Craigslist in exchange for entry into a raffle for a \$100 Amazon.com gift card. Four hundred ninety-four (197 male, mean age = 26.5 years) of these individuals were invited to take part in a follow-up study in the laboratory based on their psychic belief scores and prescreening qualifications. Forty-two believers (21 male, mean age = 27.3 years) and 42 skeptics (18 male, mean age = 26.5 years) came into the laboratory in exchange for \$35. These samples did not differ in age, $t(82) = .76, p > .25$, years of education, $t(82) = 1.41, p = .16$, or sex, $\chi^2 = 0.43, p > .25$.

Study 2 sample

In this study, 807 (444 male, mean age = 32.5 years before exclusion) individuals were initially screened and recruited on Amazon Mechanical Turk in exchange for \$0.50 and the opportunity to potentially participate in a two-part online follow-up study for additional compensation in the following weeks (\$9.50). In this screening, in addition to providing all of the

information used in Study 1 (i.e., psychic beliefs, demographics, and life satisfaction), participants reported their average high school grades (from A+ to F) in three types of classes: math, English, and creative arts. Individuals also completed the initial baseline phases of two tasks (described below).

Approximately 431 (110 believers) qualified for the study (i.e., believers and skeptics who could commit to all 3 days). Of these individuals, 166 people (88 male, mean age = 26.8 years) matched on age, gender, and years of education, were invited to complete the follow-up study across two waves of recruitment. To achieve this match, qualified believers and skeptics were selected for recruitment on the basis of these three characteristics so that both groups' distributions were similar. Of these individuals, 56 believers (27 males, mean age = 27.04) and 59 skeptics (32 males, mean age = 27.44) completed at least some part of the follow-up study. These samples did not differ in age, $t(113) = .50, p > .25$, years of education, $t(113) = .40, p > .25$, self-reported GPA, $t(113) = .80, p > .25$, or sex, $\chi^2 = 0.42, p > .25$. Some participants failed to complete the online tasks in this study and in Study 3, so for clarity we report the specific sample size in the context of each study.

Study 3 sample

In this study, 1,003 (522 male, mean age = 33.9 years before exclusion) individuals were screened and recruited on Amazon Mechanical Turk using the same procedures as in Study 2. The initial screening process included identical measures as in Study 2, and participants also filled out the DES-C (Wright & Loftus, 1999).

Approximately 402 (157 believers) qualified for the study (i.e., believers and skeptics who could commit to all 3 days). Of these individuals, 152 people (83 male, mean age = 27.0 years) matched on age, gender, years of education, and self-reported GPA were invited to complete the follow-up study across two waves of recruitment. To achieve this match, qualified believers and skeptics were selected for recruitment on the basis of these three characteristics so that both groups' distributions were similar. Of these individuals, 47 believers (22 male, mean age = 27.4 years) and 48 skeptics (27 male, mean age = 27.0 years) completed at least part of the follow-up study. These samples did not differ in age, $t(93) = .46, p > .25$, self-reported GPA, $t(93) = .38, p > .25$, or sex, $\chi^2 = 0.85, p > .25$. Although the follow-up samples in Study 3 were recruited to be matched on years of education, of the individuals who followed through and completed the study, the skeptics had more years of education than the believers (skeptical mean = 15.8 years, believer mean = 14.6 years), $t(93) = 2.37, p = .02$. To address this in a supplementary analysis, data from seven believers and eight skeptics (so that $n = 40$ for each) were removed from the sample to match the groups on years of education, and statistical analyses of these groups yielded the same results and conclusions as reported below.

Psychic beliefs questionnaire

All participants were given the Australian Sheep-Goat Scale (ASGS; Thalbourne & Delin, 1993) to measure paranormal psychic belief or skepticism. The original ASGS is composed of 16 true–false questions that measure one's belief in various psychic phenomena and is highly reliable (Cronbach's $\alpha = 0.91$ in Storm & Thalbourne, 2005). It contains three categories of items: belief in extrasensory perception (ESP, the ability to predict the future), psychokinesis (the ability to move objects using the power of one's mind), and telepathy (the ability to read another's thoughts or feelings). Sample items include “I believe that it is possible to gain information about the future before it happens, in ways that do not depend on rational prediction or normal sensory channels” and “I have had at least one vision that was not a hallucination and from which I received information that I could not have otherwise gained at that time and place.” In order to increase sensitivity, we modified the scale such that participants rated the degree of their belief for each item on a scale from 1 (no belief) to 6 (very ardent belief). Because our interest was in beliefs about

paranormal or supernatural psychic phenomena in general as opposed to idiosyncratic interpretations of some of the items or categories of items, we calculated a single score that averaged each individual's responses across all 16 items on the scale. Based on the data from the first 219 participants in Study 1, participants who were within the top third of ASGS scores (a score greater than or equal to 3.875) were classified as “believers,” and participants who were within the bottom third of ASGS scores (a score less than or equal to 2.875) were classified as “skeptics.” We used these same cutoffs for all of our studies in order to keep our definition of “believers” and “skeptics” consistent. Cronbach's α for this scale with the original 192 participants on the scale was 0.90. Cronbach's α for all 2541 participants who completed the ASGS was 0.95.

In addition to these specific paranormal belief questions, we asked participants to describe their reasons for their belief or skepticism about psychic phenomena. This question used an open-ended free response format, allowing participants to respond however they chose. For the believers in psychic phenomena that were ultimately included in our studies, the most frequently cited reasons to believe were frequent feelings of déjà vu, personal dreams and experiences that seemed psychic, or having shared paranormal beliefs with family or friends. For the skeptics that were ultimately included, they most frequently cited lack of scientific evidence or lack of personal psychic experiences as driving their belief that such phenomena do not exist.

Study 1 measures

For Study 1, participants were recruited online, and follow-up testing occurred in a single session in the lab. We assessed episodic memory accuracy and distortion using a modified version of the DRM recall task (Roediger & McDermott, 1995) and the criterial recollection task (CRT; Gallo, McDonough, & Scimeca, 2010). Analytical thinking was assessed using the *Shipley Institute of Living Scale* (Zachary, 1986). We also gave questionnaires to assess absorption (Tellegen & Atkinson, 1974), dissociative experiences (Wright & Loftus, 1999), and need for cognition (Cacioppo et al., 1984).

DRM recall task This task allowed us to assess false memory for nonstudied words under conditions where participants were motivated to try to avoid false memories. The study materials were 24 DRM lists of 15 words each that elicited high levels of false recall in the Stadler, Roediger, and McDermott (1999) norms. Each list was presented as a group of auditory stimuli presented by a speaker on the computer (each word presented for 1–1.5 s, with 250 ms interstimulus interval). Participants were informed that they would encounter several word lists and were asked to try to remember as

many words from the lists as possible. For the first 12 word lists, participants were given no warning that the words would be related to critical lures, as in the original task in Roediger and McDermott (1995). After studying and recalling these first 12 word lists, participants were forewarned that each list was associated to a word that was not actually in the list, and they should avoid falsely recalling these critical lures (as in Gallo, Roberts, & Seamon, 1997). To illustrate this relationship, they were given an example of a DRM list and its critical lure, and during the recall tests, participants were given a special box in which to write down the critical lure if they could identify it (Carneiro, Garcia-Marques, Fernandez, & Albuquerque, 2014). After studying each list, participants were given 1 minute to write down as many words as they could remember and attempt to identify the critical lure. This task yielded three critical measures: true recall of studied words (both sets of lists), false recall of critical words (both sets of lists), and the ability of participants to correctly identify the missing critical words (only on the second set of lists). We also administered a recognition memory test following the recall of all of the lists, but because the results from this test tracked the pattern of group effects found on the recall tests, we only report the recall results here for simplicity.

Criteria! recall task (CRT) This task assesses the accuracy with which people recollect information that varies in distinctiveness (Gallo et al., 2010). Stimuli were 360 easily recognizable colored pictures (e.g., umbrella, dragon, socks) and their corresponding verbal labels presented in large red font. Pictures were colored images collected from various Internet sources, with the object cropped from surrounding context. All pictures or words were presented with a white background on the computer screen. The study phase of the CRT was divided into two blocks: studying red words and studying pictures, the order of which was counterbalanced across participants. During the red-word study block, individuals were presented with a word in black font for 500 ms, and after a 100 ms delay, the same word in larger red font for 1,200 ms. After seeing the red word, in order to retain attention, participants' were asked to respond (Y/N) about whether the given item could be made in a factory. During the picture study block, individuals were instead presented with a word in black font for 500 ms, and after a 100 ms delay, a picture representing that word for 1,200 ms. In order to maintain participants' attention during this task, they were asked to judge whether the image was a highly detailed representation of the word. There were 150 trials total within each study block, and a 150 ms interstimulus interval between trials. Half of the items were studied in both formats (i.e., they were presented as both a picture and a red word), whereas the other items were presented only in one of the formats.

Each participant took three different recollection tests. On each test, the retrieval cues were black words corresponding to

items from each study format (red words and pictures) intermixed with nonstudied items. The tests differed in the retrieval instructions. On the picture test, participants were instructed to respond "yes" if they recollected that the test item had been studied as a picture (i.e., items that were only studied as a picture and items that were studied in both formats), otherwise "no" (i.e., items that were only studied as a red font or nonstudied items). On the red-word test, participants were instructed to respond "yes" if they recollected that the test item had been studied in red font (i.e., items that were only studied in red font and items that were studied in both formats), otherwise "no" (i.e., items that were only studied as a picture or nonstudied items). On these tests, because some test items had been studied in both formats, participants were told to focus only on recollecting the criterial information (e.g., red words or pictures). The exclusion test was similar to the red-word test except we did not include items studied in both formats so that if participants recollected a picture they could be sure that the item was not studied in a red font. Thus, whereas the red-word test and picture test allowed us to selectively assess recollection accuracy for words and pictures, respectively, the exclusion test assessed recollection accuracy under conditions where an exclusion rule could be employed.

Tellegen absorption scale (TAS) Absorption is defined as one's openness to experiences that are absorbing or self-altering (Tellegen & Atkinson, 1974). The original scale consists of 34 true-false items in which participants rate how often a particular scenario measuring absorption applies to them. We used a modified version of the scale (Nadon, Hoyt, Register, & Kihlstrom, 1991), ranging on a scale of 0 to 3, where 0 means *never*, 1 means *rarely*, 2 means *sometimes*, and 3 means *often*. An example of an item from the TAS is "If I wish I can imagine (or daydream) some things so vividly that they hold my attention as a good movie or story does." Individual scores were calculated as an average sum of the ratings on each item. This scale has been found to have high levels of internal reliability ($r = .88$) and high test-retest reliability ($r = .91$; Tellegen, 1982). This version of the scale was highly reliable, with a Cronbach's α of 0.95 in our study.

Dissociative experiences scale (DES) The DES-C measures the extent to which people believe they experience dissociativity in comparison to others, and it is highly reliable ($\alpha = 0.93$; Wright & Loftus, 1999). Dissociativity can be defined as the occasional failure to integrate experiences into consciousness, similar to absent-mindedness. On the DES-C, participants make a rating between 1 and 10 as to how often they feel they have particular dissociative experiences in comparison to other people. An example of an item from the DES-C is "Some people find that sometimes they are listening to someone talk and they suddenly realize that they did not hear part or all of what was said." The scale was highly reliable in

our study (Cronbach's $\alpha = 0.93$). Scores for each participant were computed by averaging across all individual items.

Shipley institute of living scale This scale has two parts: the first part is a vocabulary test, whereas the second part is an analytical test that measures reasoning or logic (Zachary, 1986). The vocabulary test contains 40 multiple-choice questions. For example, if given the word *orifice*, to which the possible choices are *brush*, *hole*, *building*, and *lute*, participants must select the word that most closely corresponds to the meaning (*hole*). The logic test contains 20 fill-in-the-blank questions in which participants must complete a sequence of letters, numbers, or words. For example, given the words “escape, scape, cape, ___,” participants must fill in the three blank spaces with letters that complete the pattern (“ape”). Scores were computed by summing the total number of correct responses on each part of the task.

Need for cognition scale This 18-item scale was included to determine whether there were differences between believers and skeptics in how much they willingly engage in difficult cognitive tasks. It is a scale from 1 to 5 in which participants indicate how “characteristic” a particular statement about cognitive motivations is of them. For example, one item reads, “I would prefer complex to simple problems.” The scale was reliable in our study (Cronbach's $\alpha = 0.88$). Scores were computed by averaging ratings for each participant.

Study 2 measures

For Study 2, participants were recruited online and then took two follow-up online testing sessions with each of the three sessions separated by a week. Online tasks were administered using the Inquisit Web Software Package 4.0.5 (Millisecond Software LLC., Seattle, WA) and Qualtrics (Qualtrics, Provo, UT). Memory distortion was assessed using an imagination inflation task. Analytical thinking was assessed using an argument evaluation task, the remote associations tests, and a conspiracy questionnaire. Working memory was assessed using two different span tasks.

Imagination inflation task (IIT) This task is designed to create autobiographical memory bias or distortion by using guided mental imagery (Garry et al., 1996). Participants were presented with 16 different events and were asked to rate the likelihood of the event having occurred during childhood on a scale of 1 (*definitely did not happen*) to 8 (*definitely did happen*). Examples of these events include “Trip and smashed hand through a window” and “Found some keys that one of my parents lost.” One week later, participants were taken through guided imagery for eight of these events. For example, “imagine that it's after school and you are playing in the house. You hear a strange noise outside, so you run to the

window to see what made the noise. As you are running, your feet catch on something and you trip and fall. Next, imagine that as you're falling you reach out to catch yourself and your hand goes through the window. As the window breaks you get cut and there's some blood.”

Participants were asked a few questions about each scenario to make sure they were actually trying to imagine the situation during the imagery task, such as “What did you trip on?” One week after the imagery, participants were again asked to rate the likelihood that the event happened in their childhood. Two scores were calculated for each participant: the average difference between the event likelihood ratings from the final day and the initial day for imagined items, and also for control items that were not imagined.

Argument evaluation task (AET) This task is designed to examine individual differences in critical thinking, or reasoning about the quality of an argument's structure while avoiding one's personal beliefs about the issue under argument (Stanovich & West, 1997). The task consists of two phases. In the first phase, participants rate their agreement with 15 different propositions about social and political issues on a scale consisting of *strongly disagree*, *disagree*, *agree*, and *strongly agree*. For example, one proposition was “Capital punishment should be outlawed.” One week later, participants read a passage about Dale, a fictitious individual, whose arguments they were instructed to evaluate. For each item, Dale stated a belief about a particular issue. These statements were identical to the items rated in the first phase. After this, he provided justification for this belief (e.g. “Capital punishment should be outlawed because killing is wrong and the moral costs of sentencing an innocent person to death are too great.”). A critic then presents an argument to counter this justification (e.g. “The prison system is very overcrowded, and it costs the state over \$25,000 per prisoner each year to maintain each prisoner.”). Finally, Dale attempts to rebut this argument with a counterargument (e.g. “The cost to the state of processing a capital punishment case through to completion averages about 10-million dollars in court and legal costs for each case.”). Participants were told to evaluate the strength of Dale's rebuttal to the counterargument on a scale of 1 to 4 (*very weak*, *weak*, *strong*, and *very strong*).

In order to evaluate each participant's performance on the AET, we used the method described in the original Stanovich and West (1997) study. Specifically, for each participant, responses across the 15 test items served as the criterion variable (i.e., the participant's final assessment of the quality of each of Dale's 15 rebuttals), which was regressed simultaneously on (a) the participant's prior belief scores for each of the 15 items (reported in the first phase of the study) and (b) the actual quality of Dale's 15 rebuttals (established by a panel of experts in the Stanovich and West, 1997, study). This procedure generated two beta weights for each participant: one that

quantified the strength of the correlation between the participant's ratings of argument quality and actual quality according to the panel of experts, independent of prior beliefs, and a beta weight that quantified the strength of the correlation between the participant's ratings of argument quality and their prior beliefs, independent of expert ratings. The former beta weight was the dependent variable that was used for group comparisons.

Working memory tasks Working memory was measured using the reading span (RSPAN) and operation span (OSPAN) tasks (Daneman & Carpenter, 1980; Turner & Engle, 1989). Participants received five sets for each span task, with each set containing seven to-be-remembered items. For each set of the RSPAN, participants were required to read a series of grammatically correct sentences and at the end of each sentence make a judgment as to whether or not the sentence was valid. An example of a valid sentence was “The seventh graders had to build a volcano for their science class,” whereas an example of an invalid sentence was “During the week of final spaghetti, I felt like I was losing my mind.” For each set of the OSPAN, participants were instead given a mathematical equation, and asked to indicate whether or not the equation was valid. An example of a valid equation was “ $(7 * 2) + 1 = 15$,” whereas an example of an invalid equation was “ $(8 * 2) - 5 = 15$.” After each sentence or mathematical validity judgment, a to-be-remembered target word was flashed on the screen for 1 second. At the end of each set of seven items, participants were asked to recall as many of these target words as possible in order. It should be noted that while seven items is a larger number than that found typically in working memory tasks, we chose this to avoid ceiling effects in our participants' performance on this task. We were not concerned about floor effects, and for the most part, participants performed very well even with these challenging trials. Two scores were calculated for each participant: the average number of words correctly recalled in the RSPAN and the average number of words correctly recalled in the OSPAN. The percentage of trials in which participants correctly identified whether the item was valid was used as a manipulation check to make sure participants were complying with the instructions of the task.

Conspiracy questionnaire The conspiracy questionnaire was intended to assess the degree to which participants endorse various conspiracy theories. It consisted of 20 items, 10 of which were conspiracy theories taken from Oliver and Wood (2014) and 10 of which were created to act as nonconspiracy controls. Whereas conspiracy items represented controversial theories of public events or situations that (by definition) are generally believed to have little substantial evidence, the control items were about noncontroversial public events or situations that are generally accepted as fact. An example of a

conspiracy item is “President Barack Obama was not really born in the United States and does not have an authentic Hawaiian birth certificate.” An example of a control item is “President Bill Clinton was impeached by the House of Representatives based on perjury and obstruction of justice but was subsequently acquitted by the Senate.” Participants were asked (1) whether or not they had heard of the claim before and (2) how strongly they agreed with each theory on a -2 to 2 scale, from *strongly disagree* to *strongly agree*. Two scores were calculated for each participant: the average agreement rating for conspiracy items and the average rating for control items.

Remote associations task (RAT) In this task, participants were given a triad of words that are closely related to a fourth word and were asked to identify the fourth word. One example was *falling, actor, dust* (solution: *star*). Scores were computed by calculating the percentage of items correctly solved. The version of the RAT used in this study consisted of 18 items (items taken from Bowers, Regehr, Balthazard, & Parker, 1990; Mednick, 1962; Mednick & Mednick 1967; see Shames, 1994).

Study 3 measures

Study 3 took place in three online sessions, similar to Study 2. This study included measures that had yielded significant group differences in the prior studies in order to determine the extent that these measures would independently replicate. These measures included the warning version of the DRM task, the working memory tasks, the argument evaluation task, the *Shipley Institute of Living Scale*, the remote associations task, and the conspiracy questionnaire. We also included the imagination inflation task, because although this task did not yield group difference in Study 2, we wanted to counterbalance the imagine and control items (see below). As in Study 1, we also assessed dissociative experiences.

The procedure for the cognitive tasks was similar to Study 1, with the following changes. First, because the DRM task was presented online in Study 3, each study list was presented visually (each word presented for 1 s, with a 1.5-s interstimulus interval). Instead of 1 min to recall each list of words participants had 45 s. For the imagination inflation task in Study 3, the imagined and unimagined items were reversed from Study 2 such that the unimagined items in Study 2 were now the imagined items in this study. This switch was intended to counterbalance the items in this task across these two studies.

Finally, in Study 3 we asked participants a few questions about Darwin's theory of evolution, including, “Although the details are still being discovered, it is an established fact that humans evolved from earlier life forms, in a way similar to natural selection described by Charles Darwin,” “If true, then

the theory of evolution makes it very unlikely that the potential for kindness or goodness to strangers is innate or genetically programed,” “A very strong belief in evolution necessarily leaves a person feeling empty or hopeless about the fate of humanity,” “There are certain things science cannot fully explain, like the existence of consciousness, or why the universe exists in the first place,” and “Overall, I am satisfied with my beliefs in how the universe works.” We also gave a question about life satisfaction: “Overall, I am satisfied with my life and my future potential.”

Results

Results from the various measures are organized by the hypothesis to which they are most relevant. Based on the cognitive differences hypothesis and related findings in the literature, our a priori predictions were that skeptics would outperform believers on the measures of memory distortion and analytical thinking but that believers would have higher dissociation, absorption, and life satisfaction scores. All tests of statistical differences reported below used the cutoff of $p < .05$ (two-tailed), except where otherwise noted. Because p values do not provide evidence in favor of the null hypothesis, we also computed the Bayesian information criteria (BIC) for each major comparison; a probability $pBIC(H1|D)$ of .50–.75 is considered as weak evidence in favor of alternative hypothesis, .75–.95 is positive evidence, .95 to .99 is strong evidence, and $>.99$ is considered very strong evidence (Masson, 2011). Furthermore, the inverse of this value is $pBIC(H0|D)$, a measure of probabilistic evidence for the null hypothesis. Correlation matrices for key measures from Studies 1, 2, and 3 can be found in the [Appendix](#).

Memory distortion measures

DRM false memory task The DRM task was administered in Study 1 ($n = 42$ in each group) and Study 3 (believer $n = 47$, skeptic $n = 48$). As can be seen from the data in Fig. 2, participants correctly recalled studied words more often than they falsely recalled nonstudied associates (i.e., critical lures), but there also was reliable false recall of critical lures even in conditions where participants were warned to avoid false recall. In contrast, false recall of noncritical words was overall very low. These patterns are consistent with previous literature using this task (see Gallo, 2006). As expected, warned participants also were able to correctly identify some but not all of the critical nonpresented words, similar to Carneiro et al. (2014).

With respect to group differences, Study 1 found several differences, but the only difference to be replicated in Study 3 was the accuracy with which individuals identified the nonstudied critical lures. In Study 1, for the first 12 word lists

(with no warning), skeptics remembered more studied words than believers, $t(82) = 3.23$, $p = .001$, $SEM = .02$, $d = .73$, $pBIC(H1|d) = .96$, but there was no difference in critical lures falsely recalled, $t(82) = .32$, $p = .74$, $SEM = .05$, $d = .07$, $pBIC(H1|d) = .10$ or false recall of noncritical words, $t(82) = -.82$, $p = .41$, $SEM = .12$, $d = .18$, $pBIC(H1|d) = .13$. For the word lists with the warning, skeptics correctly recalled more studied words than believers, $t(82) = 2.88$, $d = .63$, $SEM = .02$, $p = .005$, $pBIC(H1|d) = .86$, and fewer critical lures, $t(82) = 2.50$, $p = .01$, $SEM = .04$, $d = .55$, $pBIC(H1|d) = .71$, with no difference in falsely recalled noncritical words, $t(82) = 1.12$, $p = .24$, $SEM = .18$, $d = .26$, $pBIC(H1|d) = .18$. Skeptics were also better than believers at successfully identifying the critical lure as a missing item, $t(82) = 2.54$, $p = .01$, $SEM = .07$, $d = .55$, $pBIC(H1|d) = .72$. In Study 3 there were no significant group differences in the recall of studied words, $t(93) = 1.02$, $p = .31$, $SEM = .03$, $d = .21$, $pBIC(H1|d) = .15$, false recall of critical lures, $t(93) = 1.09$, $p = .28$, $SEM = .03$, $d = .22$, $pBIC(H1|d) = .16$, or false recall of noncritical words, $t(93) = .52$, $p = .61$, $SEM = .61$, $d = .10$, $pBIC(H1|d) = .10$. However, skeptics identified a higher percentage of critical words than did believers, $t(93) = 3.68$, $p < .001$, $SEM = .06$, $d = .75$, $pBIC(H1|d) = .98$, replicating the finding from Study 1.

Considered as a whole, we did not consistently find group differences in true recall or false recall across the two DRM studies, and the only replicated group difference was in identifying the critical missing word. While this latter measure may reflect a memory difference, it also might reflect a difference in analytical thinking (i.e., effectiveness at determining the missing word). As discussed later, the analytical thinking interpretation is more consistent with the results from our other cognitive tasks.

Criterial recollection task Results from the criterial recollection task are presented in Fig. 3. The criterial recollection task was administered in Study 1 ($n = 42$ in each group). Recollection accuracy was calculated as the ability to discriminate between criterial targets (i.e., items that had been studied in the criterial format) and noncriterial lures (i.e., items that had been studied in the noncriterial format) on each of the three tests. Collapsing across participant groups, the results replicated prior work with this task (Gallo et al., 2010). Participants were significantly less accurate on the red-word test (mean discrimination score = .29) compared to the picture test (mean discrimination score = .51), $t(84) = 7.45$, $p < .001$, $SEM = .03$, $d = .89$, demonstrating an effect of stimulus distinctiveness on recollection accuracy. Participants also were less accurate on the red-word test than on the exclusion test (mean discrimination score: .41), $t(84) = 4.52$, $p < .001$, $SEM = .03$, $d = .44$, demonstrating the benefit of the exclusion rule on recollection accuracy.

With respect to group differences, a direct comparison of the two groups on each test revealed no accuracy differences

Fig. 2 Recall DRM performance for believers and skeptics in Study 1 (laboratory) and Study 3 (online). For unrelated lures in Study 1, for the no warning lists, believers falsely recalled an average of 0.68 ± 0.12 , whereas skeptics falsely recalled 0.56 ± 0.08 . For the warning lists,

believers falsely recalled an average of 0.69 ± 0.16 per list, whereas skeptics falsely recalled 0.48 ± 0.07 per list. In Study 3, believers falsely recalled an average of 0.20 ± 0.03 per list, and skeptics falsely recalled 0.19 ± 0.03 per list. Error bars represent standard error of the mean

Fig. 3 Critical recollection task discrimination scores for believers and skeptics in Study 1 (laboratory). Discrimination scores were computed by subtracting false alarms to noncritical lures (i.e., picture words on the red-word test) from the hits to critical targets (i.e., red words on the red-word test). In addition, we also computed source confusion scores, in which there were also no group differences (red-word test–believers: 0.23 ± 0.03 , skeptics: 0.23 ± 0.04 ; picture test–believers: 0.11 ± 0.02 , skeptics: 0.11 ± 0.02 ; exclusion test–believers: 0.15 ± 0.04 , skeptics: 0.10 ± 0.04). Error bars represent standard error of the mean

in the red-word test, $t(82) = 0.14$, $p = .89$, $SEM = .05$, $d = .03$, $pBIC(H1|d) = .10$, picture test, $t(82) = 0.79$, $p = .43$, $SEM = .05$, $d = .17$, $pBIC(H1|d) = .14$, or the exclusion test, $t(82) = 1.46$, $p = .15$, $SEM = .06$, $d = .32$, $pBIC(H1|d) = .24$. We also analyzed recollection confusion scores on each test, which were calculated as the difference in false alarms to noncritical lures (i.e., items that had been studied in the noncritical format) and nonstudied lures on each test. These analyses again revealed no significant group differences on any of the three tests (all t s < 1). Overall, there was little evidence for group differences on the various measures of recollection accuracy from the CRT.

Imagination inflation task Results from the imagination inflation task are presented in Fig. 4. This task was administered in Study 2 and Study 3. Because control and imagined items were counterbalanced across these two studies, we present data collapsed across the studies here, although analysis of each separate study revealed an identical pattern of results. Data are from the 92 believers and 85 skeptics that completed all 3 days of the task. Our main dependent variable was the

Fig. 4 Imagination inflation task scores collapsed across Studies 2 and 3. Values represent the average difference between Day 3 and Day 1 ratings of event likelihood for control items (no imagination) or imagined items (imagery generated between Day 1 baseline and Day 3 ratings). Error bars represent standard error of the mean

change in ratings of event likelihood from baseline (Day 1) to follow-up (Day 3), calculated separately for control items (which were not imagined postbaseline) and imagined items (which were imagined postbaseline).

To analyze group differences on this task, we conducted a 2 (item type: control change vs. imagined change) $\times$ 2 (group: skeptic vs. believer) mixed ANOVA on these data. This analysis yielded an effect of item type: $F(1, 175) = 4.86, p = .03, MSE = .55, \eta_p^2 = .03$, reflecting the finding that estimates of event likelihood decreased from baseline (Day 1) to follow-up (Day 3) for control items, whereas imagined items tended to show no change or increases across days. An increase in these ratings across days would be expected if imagination increased confidence that the events had happened (Garry et al., 1996). There was no significant effect of group, $F(1, 175) = 1.88, p = .17, MSE = 1.15, \eta_p^2 = .01$, and no interaction, $F(1, 175) = 0.59, p = .44, MSE = .55, \eta_p^2 = .003$. The $pBIC(H1|d)$ of this interaction was .09, supporting the idea that there is no difference in the size of the inflation effects between groups. As with the other two measures, there was little evidence for group differences in this measure of memory distortion.

Working memory measures

Reading span Reading span was administered in Study 2 ($n = 59$ skeptics, 56 believers) and Study 3 ($n = 48$ skeptics, 47 believers). There were no group differences in the sentence-verification component of this task in both studies, which showed very high accuracy (over 90 %) and indicates that both groups were focused on the task ($p > .30$). Contrary to the idea that skeptics would outperform believers, results indicated that believers remembered more words per list in the correct serial position than skeptics in Study 2 (skeptic mean = 4.30, believer mean

= 5.02), $t(113) = -2.21, p = .03, SEM = .32, d = .41, pBIC(H1|d) = .51$. The group difference was not significant in Study 3 (skeptic mean = 4.44, believer mean = 5.05), $t(93) = -1.54, p = .13, SEM = .39, d = .32, pBIC(H1|d) = .26$. Pooling the data across these experiments, believers outperformed skeptics, $t(208) = -2.67, p = .008, SEM = .25, d = .37, pBIC(H1|d) = .72$. Thus, believers did not have worse working memory than skeptics, and if anything, they tended to outperform the skeptics on this measure.

Operation span Operation span was administered alongside reading span. There were no group differences in the operation-verification component of this task, which showed very high accuracy (over 90 %) and indicates that both groups were focused on the task ($p > .90$). As with reading span, believers remembered numerically more words than skeptics in Study 2 (skeptic mean = 4.77, believer mean = 5.23), $t(113) = -1.15, p = .16, SEM = .32, d = .26, pBIC(H1|d) = .20$, and Study 3 (skeptic mean = 4.93, believer mean = 5.10), $t(93) = -0.43, p = .66, SEM = .41, d = .09, pBIC(H1|d) = .10$, although neither group difference was significant. Pooling the data across these two experiments revealed no differences between believers and skeptics, $t(208) = -1.31, p = .19, SEM = .25, d = .18, pBIC(H1|d) = .14$. Thus, as with reading span, there was no evidence that skeptics outperformed believers on this working memory measure.¹

Analytical thinking measures

Shipley institute of living scale The Shipley scale was administered in Study 1 ($n = 41$ skeptics and 42 believers, as one skeptic failed to complete the task) and Study 3 ($n = 42$ skeptics and 45 believers, as six skeptics and three believers failed to complete this task), and data are presented in Fig. 5. With respect to the logic test, in Study 1 skeptics solved a significantly greater proportion of problems than did believers (skeptic mean = .78, believer mean = .67), $t(81) = 2.56, p = .01, SEM = .04, d = .56, pBIC(H1|d) = .74$. In Study 3, skeptics performed numerically higher than believers, but this difference failed to reach significance (skeptic mean = .78, believer mean = .73), $t(85) = 1.65, p = .10, SEM = .03, d = .32, pBIC(H1|d) = .29$. Pooling the logic data across the two studies, skeptics outperformed believers, $t(168) = 3.03, p < .01, d = .46, pBIC(H1|d) = .88$.

¹ As noted by a reviewer, these working memory measures might have been especially susceptible to cheating, and we were unable to determine if online participants had written down the items instead of keeping them in mind (as instructed). However, we have little reason to believe that one group might have cheated more than the other.

Fig. 5 Shipley Institute of Living Logic and Vocabulary scale in Studies 1 and 3. Values represent the percentage of trials answered correctly on each test. Error bars represent standard error of the mean

With respect to the vocabulary test, in Study 1, skeptics identified more words than did believers (skeptical mean = .81, believer mean = .74), $t(81) = 3.17$, $p < .01$, $SEM = .02$, $d = .70$, $pBIC(H1|d) = .94$. In Study 3, skeptics performed numerically higher than believers, but this difference failed to reach significance (skeptical mean = .85, believer mean = .81), $t(85) = 1.73$, $p = .09$, $SEM = .02$, $d = .38$, $pBIC(H1|d) = .33$. Pooling the vocabulary data across the two studies, skeptics outperformed believers, $t(168) = 3.42$, $p < .01$, $d = .62$, $pBIC(H1|d) = .96$.

Remote associations test Data from the RAT were collected online in Study 2 ($n = 46$ skeptics and 49 believers, as 13 skeptics and seven believers did not complete this task) and Study 3 ($n = 45$ skeptics and 42 believers, as three skeptics and five believers failed to complete this task). In Study 2, there was a trend such that skeptics identified more words than believers (skeptical mean = .48, believer mean = .40), $t(93) = 1.88$, $p = .06$, $SEM = .04$, $d = .39$, $pBIC(H1|d) = .38$. In Study 3, skeptics were numerically higher than believers (skeptical mean = .48, believer mean = .44), but this effect was not significant, $t(85) = .74$, $p = .46$, $SEM = .05$, $d = .16$, $pBIC(H1|d) = .12$. Pooling together these two studies yielded a combined sample of believers ($n = 91$) and skeptics ($n = 91$). In this combined sample, there again was a trend that skeptics identified more

words than did believers, $t(180) = 1.84$, $p = .07$, $SEM = .03$, $d = .27$, $pBIC(H1|d) = .29$. Note that although our pooled sample yielded a marginal group difference and the effect size was small to medium, the Bayesian statistics indicate that there was in fact little support for group differences on this task.

Argument evaluation task Data from the AET were collected online in both Study 2 ($n = 48$ skeptics and 45 believers, 11 skeptics and 11 believers did not complete the task) and Study 3 ($n = 48$ skeptics and 47 believers). In general, participants were quite good at evaluating the quality of the hypothetical arguments independent of their prior beliefs, as the average beta weight (representing the convergence between participant and expert evaluations, controlling for each participant's personal beliefs) was 0.29 ± 0.02 , which was significantly different than 0, $t(187) = 11.97$, $p < .001$, and comparable to the original AET results found in Stanovich and West (1997). In Study 2, although skeptics were numerically better than believers (skeptical beta: .35, believer beta: .27), this difference was not significant, $t(91) = 1.19$, $p = .24$, $SEM = .07$, $d = .25$, $pBIC(H1|d) = .17$. In Study 3, a similar pattern emerged (skeptical beta = .33, believer beta = .23), with no difference between believers and skeptics, $t(93) = 1.52$, $p = .13$, $SEM = .07$, $d = .30$, $pBIC(H1|d) = .25$. However, by combining these two samples across studies, skeptics had marginally higher betas than believers on this task, $t(186) = 1.92$, $p = .06$, $SEM = .05$, $d = .28$, $pBIC(H1|d) = .32$. In other words, skeptics were better than believers at evaluating the quality of the arguments presented, and this result was independent of prior belief. This finding is consistent with studies investigating beliefs in more general paranormal phenomena (Stanovich & West, 1998; Svedholm & Lindeman, 2012). Note that, as with the RAT, although our pooled sample yielded a marginal group difference and the effect size was small to medium, the Bayesian statistics indicate that there was in fact little support for the group differences on this task.

Conspiracy questionnaire Data from the conspiracy questionnaire (see Fig. 6) were collected in Study 2 ($n = 59$ skeptics and 56 believers) and Study 3 ($n = 48$ skeptics and 47 believers). For Study 2, a 2 (believer vs. skeptic) $\times$ 2 (conspiracy vs. control item) ANOVA revealed a main effect of item type $F(1, 113) = 532.68$, $p < .001$, $MSE = .29$, $\eta_p^2 = .83$, as all participants were more likely to endorse control items (e.g., public events generally accepted as true) than conspiracy items. Critically, there was a trend for a group effect, $F(1, 113) = 2.88$, $p = .09$, $MSE = .23$, $\eta_p^2 = .03$, and a significant interaction between item and group, $F(1, 113) = 8.64$, $p = .004$, $MSE = .29$, $\eta_p^2 = .07$, such that believers gave disproportionately higher ratings than skeptics to conspiracy items compared to control items. This pattern of results was replicated in Study 3 (main effect of item type: $F(1, 93) =$

Fig. 6 Results from the conspiracy item questionnaire. Values are on a scale of -2 to 2, where -2 is *strongly disagree* with the given item, 0 is *neither agree or disagree*, and 2 is *strongly agree*. Error bars represent standard error of the mean

471.93, $p < .001$, $MSE = .26$, $\eta_p^2 = .84$; main effect of group, $F(1, 93) = 7.11$, $p = .009$, $MSE = .23$, $\eta_p^2 = .07$, and a significant interaction, $F(1, 93) = 4.34$, $p < .001$, $MSE = .26$, $\eta_p^2 = .15$. Pooling these data together, we find a main effect of item type, $F(1, 208) = 1006.4$, $p < .001$, $MSE = 57.40$, a main effect of belief, $F(1, 208) = 9.21$, $p = .003$, $\eta_p^2 = .042$, and an interaction, $F(1, 208) = 24.13$, $p < .001$, $\eta_p^2 = .104$. The $pBIC(H1|d)$ of this interaction was larger than .99, supporting the idea that there is a very strong difference in conspiracy endorsement between believers and skeptics. These data show that believers were more likely than skeptics to endorse conspiracy theories, even though the two groups did not differ for control items, and this effect was replicated across studies.

Controlling for vocabulary differences

Interestingly, although we matched the groups on years of education and self-reported GPA, we found that skeptics outperformed believers on the vocabulary test. To the extent that vocabulary reflects the depth of one's conceptual knowledge or crystallized intelligence (due to the quality of one's educational experiences or other factors), this group difference might have been associated with some of the differences in analytical thinking that we observed. To further explore this

relationship, we simultaneously regressed our measures of analytical thinking on both psychic belief scores (binary: 1 = believer, 0 = skeptic) and Shipley Vocabulary scores. This analysis was separately done for each of the measures tapping some aspect of analytical thinking that showed a significant or near-significant group difference in the pooled data (Shipley Logic, RAT correct items, AET betas, conspiracy difference scores, and DRM critical lures identified). After controlling for individual differences in vocabulary in this way, results indicated that belief scores were not reliably related to Shipley Logic scores, $\beta = -.07$, $t = -.95$, $p = .34$, RAT scores, $\beta = .04$, $t = .42$, $p = .67$, or AET betas, $\beta = -.07$, $t = -.64$, $p = .52$. Importantly, even after controlling for vocabulary differences, belief continued to correlate with both conspiracy difference scores, $\beta = .40$, $t = 4.43$, $p < .001$, and identifying the missing DRM words, $\beta = -.23$, $t = -2.97$, $p = .003$. Thus, differences in vocabulary between believers and skeptics can explain some but not all of the group differences that we observed on the various measures thought to tap into some form of analytical thinking.

Self-report measures

Data from self-report measures are presented in Table 1.

Dissociative experiences scale-C The DES-C was administered in person in Study 1 ($n = 42$ in each group) and online in Study 3 ($n = 48$ skeptics, 47 believers). In Study 1, there was a trend such that believers reported marginally more dissociative experiences than skeptics (skeptical mean = 3.65, believer mean = 4.36), $t(82) = 1.90$, $p = .06$, $SEM = .38$, $d = .41$, $pBIC(H1|d) = .40$. Similarly, believers reported significantly more dissociative experiences than skeptics in Study 3 (skeptical mean = 2.67, believer mean = 4.68), $t(93) = 6.18$, $p < .001$, $SEM = .33$, $d = .90$, $pBIC(H1|d) > .99$. Pooling the data together, believers reported significantly more dissociative experiences than skeptics, $t(177) = 5.60$, $p < .001$, $SEM = .26$, $d = .84$, $pBIC(H1|d) > .99$. These results indicate that believers have more dissociative tendencies than skeptics, replicating Wilson and French's (2006) link between paranormal believers and dissociation and extending this link to specific beliefs about psychic phenomenon.

Tellegen absorption scale This scale was administered in Study 1 ($n = 42$ in each group). Believers had significantly higher scores than skeptics on the TAS (skeptical mean = 65.7, believer mean = 41.67), $t(82) = 5.98$, $p < .001$, $SEM = .18$, $d = 1.30$, $pBIC(H1|d) > .99$, indicating higher levels of absorption. This finding replicates Glicksohn and Barrett's (2003) link between paranormal believers and absorption, and extends this link to specific beliefs about psychic phenomenon.

Table 1 Self-Report Measures

Measure	Believers	Skeptics	
Dissociative experiences questionnaire–C			
Study 1	4.36 (.29)	3.65 (.23)	*
Study 3	4.69 (.25)	2.66 (.22)	*
Tellegen Absorption Scale	65.65 (2.9)	41.67 (2.75)	*
Need for cognition scale	2.62 (.11)	2.72 (.11)	
Life satisfaction	3.83 (.09)	3.70 (.08)	

Note. *SEM* is presented in parentheses. The DES-C score is the average rating of responses, which were on a scale of 1 to 10, where a higher score represents greater dissociativity. The TAS score is also a sum of responses, which were on a scale of 0 to 3, where a higher score represents greater absorption. The need for cognition scale score represents the average rating of responses, which were on a scale of 1 to 5, where a higher score represents greater need for cognition. The life satisfaction score represents a response to a single question about how satisfied participants were with their lives, and was on a scale of 1 to 6

Need for cognition scale Data from this scale were collected only in Study 1 ($n = 42$ in each group). There were no group differences in need for cognition (skeptical mean = 2.72, believer mean = 2.62), $t(82) = .62$, $p = .54$, $SEM = .15$, $d = .14$, $pBIC(H1|d) = .12$. The failure to find this difference is consistent with Svedholm and Lindeman (2012).

Evolution questions Believers ($n = 42$, as 5 did not complete this task) and skeptics ($n = 45$, as 3 did not complete this task) did not differ in their endorsement of Darwinian evolution as a scientific fact, with both groups endorsing evolution on average (believer mean: 6.2, skeptic mean: 5.9, out of 7), $t(85) = .71$, $p = .48$, $SEM = .33$, $d = 0.16$, $pBIC(H1|d) = .12$. Thus, although psychic believers were more likely to endorse supernatural concepts, the groups were equally likely to believe other scientific conclusions about the world. Interestingly, the groups did show differences in their understanding of the implications of evolution. For the question, “If true, then the theory of evolution makes it very unlikely that the potential for kindness or goodness to strangers is innate or genetically programmed,” believers (mean = 3.86) gave significantly higher ratings than skeptics (mean = 3.07), $t(85) = 2.25$, $p = .03$, $SEM = .35$, $d = .48$, $pBIC(H1|d) = .57$. On the question that read, “Overall, I am satisfied with my beliefs in how the universe works,” believers (mean = 5.29) gave significantly lower ratings than skeptics (mean = 5.87), $t(85) = 2.13$, $p = .04$, $SEM = .27$, $d = .45$, $pBIC(H1|d) = .51$. These data suggest that both groups believed in evolution, but that skeptics were more satisfied with the perceived implications of these positions than were believers.

Life satisfaction Answers to the question about life satisfaction came from the prescreening in all three studies. Pooling across the three studies (145 believers, 149 skeptics) there was no group difference between believers (3.83) and skeptics (3.70) in life satisfaction, $t(292) = 1.19$, $p = .26$, $SEM = .12$, $d = .13$, $pBIC(H1|d) = .10$. Because this question was administered during prescreening, we also pooled all of the prescreening participants who answered the life satisfaction question and took the ASGS ($n = 2541$ who completed the prescreening process, no exclusions). An ordinary least squares regression analysis on these data showed a significant positive correlation between the degree of psychic beliefs and life satisfaction, $\beta = .19$, $t(2,539) = 8.99$, $p < .001$, even after controlling for age, sex, and years of education. This finding replicates prior research demonstrating that paranormal beliefs are associated with greater life satisfaction than skepticism (Kennedy et al., 1994; Parra & Coretta, 2013).

General discussion

This research provided a large-scale test of the cognitive differences hypothesis, revealing that psychic believers and skeptics did not consistently differ on memory measures but did differ on measures tapping different aspects of analytical thinking. With respect to memory, we found little evidence that psychic beliefs are associated with greater memory biases and errors, even though we used three different memory tasks tapping different aspects of memory accuracy and distortion. While psychic believers did show an elevated DRM illusion in Study 1, this difference was only found in the warning condition of that study, with no difference in the no-warning condition. Moreover, these groups did not significantly differ in recollection accuracy on any of the criterial recollection tests in Study 1, and the group difference in the DRM warning task did not replicate in an independent follow-up study. There also were no group differences in the imagination inflation task for autobiographical memories, which was administered in two different studies. While it is always a possibility that the lack of group differences on such memory measures is due to insufficient statistical power, we used similar or larger sample sizes as studies that have observed significant group differences in these same kinds of tasks (e.g., Gallo, Cotel, Moore, & Schacter, 2007; Meyersburg et al., 2009).²

² To make this point more concrete, Balota et al. (1999) and Watson, McDermott, and Balota (2004) found significant differences across age groups the DRM recall task, whereas Gallo et al. (2007) and McDonough, Wong, and Gallo (2013) found significant differences across age groups on the criterial recollection task, even though each of these studies used fewer participants per comparison group than we did in the current study. Thus, the current study was sufficiently powered to detect group differences at least as large as these aging-related effects, which themselves can be quite subtle at times.

Moreover, the current studies had sufficient power to detect group differences in other measures, such as personality measures that have been linked to memory distortion (dissociation and absorption) and several of the analytical thinking tasks. As a whole, our results indicate that there is little or no relationship between individual differences in memory accuracy or distortion and psychic beliefs.

The finding that believers and skeptics did not differ in memory accuracy or susceptibility to memory distortion is consistent with Rose and Blackmore (1997, 2001), who failed to find a correlation between more general paranormal beliefs and memory errors. Our results extend this result to psychic believers, using multiple measures of memory accuracy and distortion as well as a between-groups approach that should have been maximally sensitive to any group differences. Our results also help clarify interpretation of the results reported by Clancy et al. (2002) and Meyersburg et al. (2009), who found that individuals with autobiographical memories of extraterrestrial alien contact or past lives (respectively) had higher levels of false memory on the DRM task. In the Introduction we noted that these findings suggest a link between paranormal beliefs and a propensity for false memories, but the results of our current studies do not support this interpretation. Instead, these other studies might best be interpreted as showing individual differences in the propensity for memory distortion in both autobiographical and laboratory situations, as the authors of those studies had initially concluded.

In contrast to the memory distortion hypothesis, several of our comparisons replicated the link between paranormal beliefs and analytical thinking that has been reported in the literature (e.g. Krummenacher, Mohr, Haker, & Brugger, 2010; Rogers et al., 2009; Watt & Wiseman, 2002). Overall we found evidence that skeptics outperformed believers on four measures that require some form of analytical thinking: (1) the logic subscale of the *Shipley Institute of Learning Scale*, which required them to complete patterns of words, letters, and numbers; (2) the remote associations task, which required the identification of a nonpresented word that was related to three presented words; (3) the rejection of conspiracies on the conspiracy questionnaire, suggesting that skeptics engage in more critical or analytical thinking about these kinds of conspiracies than do believers; and (4) the argument evaluation task, which requires critical thinking about the quality of arguments made during a hypothetical debate. These group differences were most robust on Shipley Logic and the conspiracy questionnaire, whereas the group differences on remote associations and the argument evaluation were only marginally significant with our conservative statistical threshold and received little support from the Bayesian statistical approach. However, given that all of these differences were in the direction predicted by the analytical thinking hypothesis, we believe that they are all meaningful. In fact, the pooled group effect sizes (Cohen's *d*) across these four measures ranged

from .27 (on remote associations) to .68 (on the conspiracy questionnaire), demonstrating small-to-medium-sized effects for the weakest relationships and large-sized effects for the strongest relationships, at least by that standard.

In addition to our measures of analytical thinking, we also found that skeptics were better than believers at identifying the nonstudied associate when presented with DRM word lists. These effects were reliable, as they were obtained in two separate studies, one conducted in the laboratory and one conducted online. The lack of consistent group differences in the other memory tasks suggests that some process other than memory accuracy, such as analytical thinking, drove this effect. In the DRM task, the ability to identify the nonstudied associate requires participants to rapidly analyze the associative relationships between the presented words and mentally generate candidate words that link them all together. This aspect of the DRM task is similar to the analytical thinking required to solve the remote associations task (e.g., identifying a missing word that links the three presented words; see Howe & Wilkinson, 2011) or the Shipley Logic Scale (analyzing patterns of words, or letters or numbers to determine the next item in the sequence). In all of these cases, believers were less likely than skeptics to analyze the evidence provided to them in order to identify critical missing pieces of information.

As a whole, our analytical thinking findings are consistent with prior work and also generalize the kinds of analytical thinking tasks that differ between believers and skeptics. The differences obtained in our study are particularly striking given that our groups were matched on years of education, age, and self-reported GPA, and also given that skeptics did not outperform believers on measures of working memory (if anything, the believers did slightly better than the skeptics on one of the span tasks). These findings point to a specific cognitive difference in analytical thinking that is independent from these other cognitive and intellectual domains. One caveat to this conclusion is that skeptics also outperformed believers on a vocabulary test, which may be considered a more sensitive measure of conceptual knowledge or crystallized intelligence than years of education or self-reported GPA. When these group differences in vocabulary were statistically controlled, we found that only the group differences in conspiracy items and identification of DRM critical items remained significant. Thus, differences in conceptual knowledge or crystallized intelligence between believers and skeptics might have been in play for some of the group differences in analytical thinking that we observed, but they cannot explain all of the group differences that we observed. Future work will need to more directly investigate the relationship between conceptual knowledge, analytical thinking, and psychic beliefs.

Related to these ideas, although paranormal beliefs have previously been linked to less belief in important scientific concepts, such as evolution (Eder et al., 2011) and medicine (Lindeman, Svedholm, Takada, Lonnqvist, & Verkaslo,

2011), we found that our groups were equally likely to endorse Darwin's theory of evolution. Thus, at least in our sample, there was no evidence that paranormal psychic beliefs were related to the rejection of scientific beliefs more generally.

Implications and future directions

We found that psychic believers were less likely than skeptics to successfully analyze and evaluate information about the world and their own experiences, even though the groups did not differ in basic memory abilities. While these findings do not demonstrate a causal link between analytical thinking and the development of paranormal psychic beliefs, they are consistent with the hypothesis that this cognitive difference may foster beliefs about the existence of psychic phenomena. Indeed, many of our psychic believers claimed to have had personal experiences with ESP (e.g., dreams that seemed to come true) when, in fact, more careful scrutiny of these experiences might have shown them to be nonpredictive of future events, on average, or to be based on recently experienced events that themselves might have been predictive of future events.

One interesting question raised by our findings is the extent that these group differences in analytical thinking were due to differences in a fundamental ability to engage in effective analytical thinking or if they instead were due to individual differences in information processing style or motivation to engage in analytically thinking. This is an important distinction because the former suggests a relatively fixed aspect of information processing, whereas the latter suggests a more flexible difference (e.g., a potential trade-off between effortful analytical thinking and faster and less effortful intuitive thinking). Some literature suggests that paranormal believers and skeptics differ in cognitive style or motivation to use effortful thought processes (Pennycook et al., 2012; Shenhav, Rand, & Greene, 2012; Stanovich & West, 2008), and group differences in motivation or information processing style could explain our analytical thinking results without appealing to group differences in cognitive ability. On the other hand, we found little group differences on cognitively demanding measures of memory distortion and working memory and also found no group differences in self-reported need for cognition (Cacioppo et al., 1984). These latter findings speak against a group difference in motivation to complete cognitive tasks in general, so that any group differences in information processing style or motivation would have to be specific to those cognitive tasks that required analytical thinking. Future work will be needed to determine the extent that these cognitive differences are due to information processing ability, information processing style, or a combination of both.

Another interesting question raised by our findings is the extent that these group differences in analytical thinking are related to other cognitive differences reported in the literature. Some research suggests that paranormal believers are more likely to identify patterns in noise or draw meaningful connections between unrelated events (Blackmore, 1992; Krummenacher et al., 2010; Rominger, Weiss, Fink, Schuler, & Papousek, 2011). For example, when presented with images composed of visual noise, Blackmore and Moore (1994) found that psychic believers were significantly more likely to see patterns in the noise than skeptics (see also Riekkari et al., 2013; van Elk, 2013). Similarly, using a word association task, Gianotti et al. (2001) found that believers in a variety of paranormal phenomena produced more idiosyncratic words or more distant associations than did skeptics. It also has been suggested that some kinds of paranormal beliefs – and particularly paranormal experiences – might be related to abnormal neurocognitive development or schizotypy (see Brugger & Graves, 1997). While we did find a link between psychic beliefs and personality characteristics such as dissociative experiences and absorption, which may relate to these other cognitive factors, we did not directly investigate these other cognitive factors in the current study. Future work should determine whether these other findings could be explained by group differences in analytical thinking or whether they represent separate cognitive factors that also might contribute to the development of paranormal beliefs.

In conclusion, our results indicate that psychic beliefs were not related to a greater propensity for episodic memory distortion or poor working memory, but psychic beliefs were related to individual differences in tasks tapping into different forms of analytical thinking as well as vocabulary scores. These findings suggest that differences in analytical thinking as well as conceptual knowledge might foster the development of psychic beliefs. It is important to reiterate, though, that noncognitive factors also are likely to play a role in the development of psychic beliefs. Approximately 70 % of the believers in our study indicated that their psychic beliefs were in line with those of close friends and family, likely representing a mix of sociocultural factors. This link suggests that being raised in a community of individuals where paranormal beliefs are accepted makes one more likely to accept them on personal level, potentially increasing one's sense of belonging and consistency in one's worldview. We also found that psychic belief was correlated with life satisfaction in our entire sample, suggesting that holding paranormal psychic beliefs can have psychological benefits for some individuals. Given these links, an important direction for future research will be to determine the extent that individual differences in these other psychological and sociocultural factors might interact with differences in analytical thinking in driving psychic beliefs.

Appendix

Table 2 Study 1 Correlation Matrix

	1	2	3	4	5	6	7	8	9	10
1. ASGS	x	.252*	-.268*	-.312**	.232*	.647**	.042	-.401**	-.256*	.347**
2. DRM False Recall	x	x	-.456**	-.607**	-.053	.101	-.116	-.340**	-.460**	.136
3. DRM Target Recall	x	x	x	.577**	.106	-.118	.067	.442**	.576**	-.097
4. DRM Lure Identity	x	x	x	x	-.003	-.101	.053	.352**	.461**	-.002
5. DES-C	x	x	x	x	x	.322**	-.069	.204	.186	-.123
6. TAS	x	x	x	x	x	x	.165	-.193	-.051	.269
7. NFC	x	x	x	x	x	x	x	.098	.256*	.361**
8. Shipley Vocabulary	x	x	x	x	x	x	x	x	.574**	-.312**
9. Shipley Logic	x	x	x	x	x	x	x	x	x	.001
10. Life Satisfaction	x	x	x	x	x	x	x	x	x	x

Note. DRM scores are from the warning DRM task. DRM False Recall represents false recall of critical lures only

* $p < .05$. ** $p < .01$

Table 3 Study 2 Correlation Matrix

	1	2	3	4	5	6	7
1. ASGS	x	.166	.064	.319**	-.033	.096	-.021
2. Reading Span Score	x	x	.781**	.040	-.043	.005	.052
3. Operation Span Score	x	x	x	-.042	-.061	-.014	.166
4. Conspiracy Score	x	x	x	x	-.043	-.113	-.096
5. AET Beta	x	x	x	x	x	-.066	-.166
6. IIT Score	x	x	x	x	x	x	-.010
7. Life Satisfaction	x	x	x	x	x	x	x

Note. Conspiracy score represents difference between conspiracy and nonconspiracy item baseline. Imagination inflation score represents difference between Day 3–Day 1 inflation for imagined items and Day 3–Day 1 inflation for control items

* $p < .05$. ** $p < .01$

Table 4 Study 3 Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. ASGS	x	.175	.058	.122	-.048	-.374**	-.158	-.179	-.178	.522**	.026	.424**	-.114
2. Reading Span Score	x	x	.865**	-.195	.420**	.024	.069	-.055	.138	.043	-.052	.200	.222*
3. Operation Span Score	x	x	x	-.213*	.434**	.125	.061	-.079	.217*	-.020	-.098	.079	.199
4. DRM False Recall	x	x	x	x	-.170	-.185	.148	-.074	.029	.098	.265*	.051	-.065
5. DRM Target Recall	x	x	x	x	x	.175	.194	.237*	.267*	-.136	-.057	-.051	.208*
6. DRM Lure Identity	x	x	x	x	x	x	.158	.291**	.165	-.169	.019	-.282**	-.054
7. AET Beta	x	x	x	x	x	x	x	-.254*	.230*	-.083	-.079	-.238*	.118
8. Shipley Vocabulary	x	x	x	x	x	x	x	x	.457**	-.034	-.019	-.426**	.042
9. Shipley Logic	x	x	x	x	x	x	x	x	x	-.022	-.013	-.236*	.123
10. DES	x	x	x	x	x	x	x	x	x	x	-.097	.022	-.299**
11. IIT Score	x	x	x	x	x	x	x	x	x	x	x	.029	-.080
12. Conspiracy Score	x	x	x	x	x	x	x	x	x	x	x	x	-.175
13. Life Satisfaction	x	x	x	x	x	x	x	x	x	x	x	x	x

Note. Conspiracy score represents difference between conspiracy and nonconspiracy item baseline. Imagination inflation score represents difference between Day 3–Day 1 inflation for imagined items and Day 3–Day 1 inflation for control items

* $p < .05$. ** $p < .01$

References

- Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., III, McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology*, 16, 361–384.
- Blackmore, S. J. (1992). Psychic experiences: psychic illusions. *Skeptical Inquirer*, 16, 367–376.
- Blackmore, S. J., & Moore, R. (1994). Seeing things: visual recognition and belief in the paranormal. *European Journal of Parapsychology*, 10, 91–103.
- Blackmore, S. J. & Rose, N. (1997). Reality and imagination: a psi-conducive confusion? *Journal of Parapsychology*, 61, 321–335.
- Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, 22, 72–110.
- Braswell, G. S., Rosengren, K. S., & Berenbaum, H. (2012). Gravity, God and ghosts? Parents' beliefs in science, religion, and the paranormal and the encouragement of beliefs in their children. *International Journal of Behavioral Development*, 36, 99–106.
- Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: the conspiracy mentality questionnaire. *Frontiers in Psychology*, 4, 1–15.
- Brugger, P., & Graves, R. E. (1997). Testing vs. believing hypotheses: magical ideation in the judgment of contingencies. *Cognitive Neuropsychiatry*, 2, 251–272.
- Cacioppo, J., Petty, R., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48, 306–307.
- Carneiro, P., Garcia-Marques, L., Fernandez, A., & Albuquerque, P. (2014). Both associative activation and thematic extraction count, but thematic false memories are more easily rejected. *Memory*, 22, 1024–1040.
- Clancy, S., McNally, R., Schacter, D., Lenzenweger, M., & Pitman, R. (2002). Memory distortion in people reporting abduction by aliens. *Journal of Abnormal Psychology*, 111, 455–461.
- Corlett, P. R., Simons, J. S., Pigott, J. S., Gardner, J. M., Murray, G. K., Krystal, J. H., & Fletcher, P. C. (2009). Illusions and delusions: relating experimentally-induced false memories to anomalous experiences and ideas. *Frontiers in Behavioral Neuroscience*, 3, 53.
- 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, 711–719.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- 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, 517–534.
- Eisen, M. L., & Lynn, S. J. (2001). Dissociation, memory and suggestibility in adults and children. *Applied Cognitive Psychology*, 15, S43–S73.
- French, C., Santomauro, J., Hamilton, V., Fox, R., & Thalbourne, M. (2008). Psychological aspects of the alien contact experience. *Cortex*, 44, 1387–1395.
- Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York: Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, 38, 833–848.
- Gallo, D. A. (2013). Retrieval expectations affect false recollection: insights from a criterial recollection task. *Current Directions in Psychological Science*, 22, 316–323.
- Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: the importance of event distinctiveness. *Psychology & Aging*, 22, 209–213.
- Gallo, D. A., McDonough, I. M., & Scimeca, J. M. (2010). Dissociating source memory decision in the prefrontal cortex: fMRI of diagnostic and disqualifying monitoring. *Journal of Cognitive Neuroscience*, 22, 955–969.
- Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remember words not presented in lists: can we avoid creating false memories? *Psychonomic Bulletin & Review*, 4, 271–276.
- Gallup. (2005). Three in four Americans believe in paranormal: Little change from similar results in 2001. Retrieved from <http://www.gallup.com/poll/16915/Three-Four-Americans-Believe-Paranormal.aspx>
- Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, 3, 208–214.
- Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, 336, 493–496.
- Gianotti, L. R. R., Mohr, C., Pizzagalli, D., Lehmann, D., & Brugger, P. (2001). Associative processing and paranormal belief. *Psychiatry and Clinical Neurosciences*, 55, 595–603.
- Glicksohn, J., & Barrett, T. R. (2003). Absorption and hallucinatory experience. *Applied Cognitive Psychology*, 17, 833–849.
- Howe, M. L., & Wilkinson, S. (2011). Using story contexts to bias children's true and false memories. *Journal of Experimental Child Psychology*, 108, 77–95.
- Irwin, H. J. (2009). *Psychology of paranormal belief: A researcher's handbook*. Hertfordshire: University of Hertfordshire Press.
- Kennedy, J. E., Kanthamani, H., & Palmer, J. (1994). Psychic and spiritual experiences, health, well-being, and meaning in life. *Journal of Parapsychology*, 58, 353–383.
- Krummenacher, P., Mohr, C., Haker, H., & Brugger, P. (2010). Dopamine, paranormal belief, and the detection of meaningful stimuli. *Journal of Cognitive Neuroscience*, 22, 1670–1681.
- Lindeman, M., Svedholm, A. M., Takada, M., Lonnqvist, J., & Verkasto, M. (2011). Core knowledge confusions among university students. *Science & Education*, 20, 439–451.
- 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.
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43, 679–690.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, 69, 220–232.
- Mednick, S. A., & Mednick, M. T. (1967). *Examiner's manual: Remote associates test*. Boston: Houghton Mifflin.
- Messer, W. S., & Griggs, R. A. (1989). Student belief and involvement in the paranormal and performance in introductory psychology. *Teaching of Psychology*, 16, 187–191.
- Meyersburg, C. A., Bogdan, R., Gallo, D. A., & McNally, R. J. (2009). False memory propensity in people reporting recovered memories of past lives. *Journal of Abnormal Psychology*, 118, 399–404.
- Musch, J., & Ehrenberg, K. (2002). Probability misjudgment, cognitive ability, and belief in the paranormal. *British Journal of Psychology*, 93, 167–177.
- Nadon, R., Hoyt, I. P., Register, P. A., & Kihlstrom, J. F. (1991). Absorption and hypnotisability: context effects reexamined. *Personality Processes and Individual Differences*, 60, 144–153.
- Oliver, E. J., & Wood, T. J. (2014). Conspiracy theories and the paranoid style(s) of mass opinion. *American Journal of Political Science*, 58, 952–966.
- Parra, A., & Corbetta, J. M. (2013). Paranormal experiences and their relationship with the meaning of life. *Liberabit*, 19, 251–2583.

- Pennycook, G., Cheryne, J. A., Sleli, P., Koehler, D., & Fugelsang, J. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, 123, 335–346.
- Riekk, 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, 150–155.
- Roberts, M. J., & Seager, P. B. (1999). Predicting belief in paranormal phenomena: a comparison of conditional and probabilistic reasoning. *Applied Cognitive Psychology*, 13, 443–450.
- Roediger, H., & McDermott, K. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 803–814.
- Rogers, P., Davis, T., & Fisk, J. (2009). Paranormal belief and susceptibility to the conjunction fallacy. *Applied Cognitive Psychology*, 23, 524–542.
- Rominger, C., Weiss, E. M., Fink, A., Schuler, G., & Papousek, I. (2011). Allusive thinking (cognitive looseness) and the propensity to perceive “meaningful” coincidences. *Personality and Individual Differences*, 51, 1002–1006.
- Rose, N., & Blackmore, S. (2001). Are false memories psi-conducive? *Journal of Parapsychology*, 65, 125–144.
- Shames, V. A. (1994). Is there such a thing as implicit problem solving? Unpublished doctoral dissertation, University of Arizona.
- Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: cognitive style influences belief in God. *Journal of Experimental Psychology: General*, 141, 423–428.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, 26, 4–7.
- Smith, M. D., Foster, C. L., & Stovin, G. (1998). Intelligence and paranormal belief: examining the role of context. *Journal of Parapsychology*, 62, 65–77.
- Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, 27, 494–500.
- 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, 342–357.
- Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, 127, 161–188.
- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94, 672–695.
- Storm, L., & Thalbourne, M. A. (2005). The effect of a change in pro attitude on paranormal performance: a pilot study using naïve and sophisticated skeptics. *Journal of Scientific Exploration*, 19, 11–29.
- Svedholm, A. M., & Lindeman, M. (2012). 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, 303–319.
- Tellegen, A. (1982). *A brief manual for the differential personality questionnaire*. Minneapolis: Department of Psychology, University of Minnesota.
- Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences (“absorption”), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, 83, 268–277.
- Thalbourne, M., & Delin, P. (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.
- Tobacyk, J., Miller, M. J., & Jones, G. (1984). Paranormal beliefs in high school students. *Psychological Reports*, 55, 255–261.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Unsworth, N., & Brewer, G. A. (2009). Examining the relationships among item recognition, source recognition, and recall from an individual differences perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1578–1585.
- van Elk, M. (2013). Paranormal believers are more prone to illusory agency detection than skeptics. *Conscious Cognition*, 22, 1041–1046.
- Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 76–85.
- Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in Deese/Roediger-McDermott paradigm: assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition*, 32, 135–141.
- Watt, C., & Wiseman, R. (2002). Experimenter differences in cognitive correlates of paranormal belief and in psi. *Journal of Parapsychology*, 66, 371–385.
- Wilson, K., & French, C. (2006). The relationship between susceptibility to false memories, dissociativity, and paranormal belief and experience. *Personality and Individual Differences*, 41, 1493–1502.
- Wiseman, R., & Watt, C. (2006). Belief in psychic ability and the misattribution hypothesis: a qualitative review. *British Journal of Psychology*, 97, 323–338.
- Wright, D., & Loftus, E. (1999). Measuring dissociation: comparison of alternative forms of the dissociative experiences scale. *The American Journal of Psychology*, 112, 497–519.
- Zachary, R. (1986). *Shipley institute of living scale: Revised manual*. Los Angeles: Western Psychological Services.
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., ... Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory*, 18, 543–555.