Belief beyond logic: An investigation of the role of intuitive reasoning in delusion formation and maintenance

By

Alycia Budd

Bachelor of Psychological Science (Honours)

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy (Psychology)

College of Education, Psychology & Social Work

18th March 2024

Contents

Summary

Clinical delusions are largely characterised by inappropriate levels of conviction for beliefs that persist in the face of contradictory evidence. These beliefs often share similar thematic content, for example, persecution, grandiosity, or control. While the precise mechanisms underlying the development of high conviction in delusional beliefs remain unclear, research has more recently focused on the role of intuitive reasoning processes in the formation and maintenance of delusions.

In tandem, there is a robust literature investigating how repetition influences truth evaluation in healthy individuals. This phenomenon, known as *the truth effect,* has shown that statements that have been encountered before are more likely to be judged as true than novel statements. The primary mechanism underlying this effect is a metacognitive experience termed *processing fluency* (i.e., the ease or difficulty with which a stimulus is processed), which is then used intuitively as a cue for truth.

Over a series of truth effect experiments, this project explored the use of processing fluency in perceptions of truth, and how it might relate to the development of erroneous conviction as seen in clinical delusions. Specifically, the investigation sought to elucidate whether the truth effect behaved differently in response to different types of thematic content, and across individual differences in trait schizotypy or delusion-proneness.

The first empirical chapter explored whether truth judgements are influenced by the thematic content being evaluated (i.e., delusion-relevant, or neutral). Participants completed a truth effect task containing a series of delusion-relevant and neutral trivia statements, of which half were repeated once during the experiment. Contrary to expectations, the truth effect was

smaller for delusion-relevant than for neutral items; however, differences in novel truth ratings between the two statement types prompted further investigation.

The second empirical chapter identified familiarity as a significant factor affecting baseline ratings in the preceding experiments. An item analysis was conducted, and a final truth effect experiment for delusion-relevant and neutral content was performed. The findings again revealed that the truth effect, if anything, was smaller for delusion-relevant items than for neutral items.

The third empirical chapter investigated whether individual differences in trait schizotypy predicted the size of the truth effect across the preceding chapters. These individual differences were also compared across the truth effect for delusion-relevant and neutral items. The findings revealed an interaction between the truth effect, delusion-relevant content, and positive trait schizotypy, whereby higher levels of positive schizotypy predicted a larger truth effect for delusion-relevant content.

The fourth and final empirical chapter investigated whether statements that are self-referential elicit a larger truth effect than content relating to another known person. These results were also examined across individual differences in trait schizotypy. However, no truth effect was observed in this instance.

This project provides a novel contribution to research surrounding the role of processing fluency in perceptions of truth among individuals who are more delusion-prone or higher in trait schizotypy, and further implicates the role of thematic content in belief formation.

Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university
2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and
3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Signed Alycia Budd

Date 26/10/2023

Acknowledgements

Dr Oren Griffiths. Thank you for your time, patience, and encouragement throughout my PhD. I am grateful for your commitment to my development as a researcher and as a writer. I have appreciated your words of reassurance, as well as your knowledge, advice, and consistent support over the years. Your dedication to the success of your students and peers is significant and does not go unnoticed.

Professor Mike Nicholls. It has been wonderful to be part of your Brain and Cognition Lab over the past several years. Thank you for your words of encouragement and advice at those pivotal moments when I needed them.

To my mentor, **Professor Marika Tiggemann.** Thank you for being a consistent source of support and guidance for me, especially during times when I was lacking confidence. I am grateful for your ability to help me reframe setbacks, and for helping me celebrate the small successes along the way.

To the love of my life, **Wynn.** I am eternally grateful for your patience, optimism, and encouragement over the years. Thank you for being there for me through difficult passages of time, and for always believing in me. You are my greatest source of strength. I am so proud of you and what we have accomplished together in these recent years. This work is dedicated to you.

To my parents, siblings, and friends, I could not have done this without your endless love and support. Thank you for making me laugh and for being there through challenging times. Laura, I could not have made it through Tuesdays without you. James, thank you for indulging in countless conversations with me about my research, and for showing such an interest in my work over the years.

To my friends in the Brain and Cognition Lab, it has been an honour getting to know you all. Sal, John, David, Steph, Tess, and Xin, thank you for your friendship and guidance throughout my PhD. Your encouragement and good humour have helped carry me through this epic challenge, and I only hope I was able to do the same for you all in return.

I would also like to express my gratitude for the financial support I received through Flinders University and the Australian Government Research Training Program. My research scholarship allowed me to dedicate time to my studies, and awards from the College of Education, Psychology and Social Work allowed me to conduct my research and present my findings at conferences. Without this financial support, I would not have been able to achieve what I have over these past few years.

Finally, thank you to all who participated in my experiments. Without you, we could not have gained this knowledge.

List of Figures

List of Tables

Chapter 1: Introduction

The scientific study and understanding of delusions as they present themselves in psychosis – though longstanding, productive and somewhat fruitful – is still far from complete. In this search for understanding, the processes of reasoning via which one comes to believe in an idea or perspective have been an important focal point of investigation. This thesis seeks to contribute to that understanding of reasoning processes through exploring the influence of, and reliance upon, intuitive reasoning in the development of conviction in both ordinary and delusional belief.

Delusions

Clinical delusions are largely characterised by inappropriate levels of conviction for beliefs that do not waver in the face of contradictory evidence (American Psychiatric Association, 2013). Indeed, as specified in the most recent Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American Psychiatric Association, 2013), holding an idea or belief with a high level of conviction is the primary feature that should be used in distinguishing delusions from strongly held beliefs. This inappropriate conviction for beliefs can lead to significant disruptions in a person's life, eliciting thoughts, emotions, and behaviours that have a negative impact on social, occupational, and personal wellbeing outcomes (Beck, 2020; Freeman et al., 2002).

The actual definition of delusions as described in the current DSM is that they are "fixed beliefs that are not amenable to change in light of conflicting evidence" (American Psychiatric Association, 2013). This definition has long been a topic of philosophical and psychological debate (Garety, 1991; Kiran & Chaudhury, 2009). Even within the field of clinical psychology, the diagnostic definition has been subject to changes, refinements, and additional criteria over the years. For example, key words such as 'false' – appearing in previous DSM definitions –

and ‘fixed’, which still appears in the current edition (but with added specifications), have been criticised. For example, some have pointed out that not all delusions can necessarily be proven false. In some cases, a person with the delusion might be correct about their belief, but only by coincidence (e.g., delusions of a spouse being unfaithful; Casey & Kelly, 2019). Others have observed that some delusional beliefs can change over time (Appelbaum, Robbins, & Vesselinov, 2004) or, through therapeutic intervention, become less fixed (Garety & Freeman, 2013). Thus, without rejecting definitions altogether, it is important to remain cognisant of the limitations of current scientific understandings while conducting research on the formation and maintenance of delusions. Nevertheless, a framework for identifying delusions for the purpose of clinical diagnosis and therapeutic intervention remains vital; we require a practical definition that allows us to differentiate between those with overvalued or “delusion-like” ideas, and those with clinically diagnosable delusions.

In recent years, there has been a gradual movement towards emphasising the *experiential aspect* of delusional beliefs in clinical diagnosis: that is, how they are experienced by the patient in, for example, frequency, preoccupation, and level of conviction. This is a movement away from a previous focus on delusional *content* in clinical diagnosis (Garety & Freeman, 2013; Kiran & Chaudhury, 2009). This shift is partly because belief content is not informative enough in the identification and classification of delusions – for example, not all bizarre beliefs are clinical delusions (e.g., in the case of some religious or paranormal beliefs), and not all delusional beliefs are bizarre (think of the paranoia regarding unfaithfulness in the case described above). In addition, such information does not predict the severity of the illness, the ideal course of treatment, or a patient’s prognosis (Tandon et al., 2009). As such, dimensional features of delusions – including conviction, frequency, distress, functional impairment, and preoccupation – have become increasingly important in diagnosis and

assessment. A benefit of using experiential descriptors that are unrelated to the type or content of a belief is that they allow clinicians to identify and diagnose delusions without requiring an analysis of the plausibility or belief subject, which may be idiosyncratic or inconsistent across time (Appelbaum, Robbins, & Vesselinov, 2004). Moreover, non-content descriptors such as conviction are increasingly being recognized as important clinical features across a broad range of diagnoses, including schizophrenia, bipolar disorder, major depression, and drug and alcohol use disorder (Appelbaum, Robbins, & Roth, 1999).

Further to this focus on the experiential features of delusions, many researchers and clinicians have adopted the *continuum model of psychosis* (van Os et al., 2009) – that is, the hypothesis that psychosis falls on a continuum, ranging from nonclinical, subthreshold states and personality traits, up to the clinically recognised and diagnosable symptoms typical of psychotic disorders. For example, trait schizotypy is a well-documented personality dimension that includes features such as magical ideation and unusual perceptual experiences, which on their own are not symptoms of psychiatric illness (Grant, 2018; Meehl, 1962). As encounters such as these become more frequent, disruptive, and distressing, a person may be at risk of psychosis, and these experiences may meet the criteria for delusions and hallucinations (Mason, 2016). Individuals on the highest end of this spectrum of experience may then receive a clinical diagnosis of a psychotic disorder. For this reason, schizophrenia is often now referred to as a spectrum disorder, rather than a dichotomous categorical 'has/has not' diagnostic criteria.

The development of delusions

Despite the prevalence of delusional beliefs among clinical populations such as schizophrenia, there remains a lack of consensus regarding the precise mechanisms underlying the formation and maintenance of delusions. Indeed, there is an extensive literature supporting

the involvement of several potentially separable factors in the development of clinical delusions, including cognitive factors such as the jumping to conclusions bias (Garety & Freeman, 1999) and the bias against disconfirmatory evidence (Woodward et al., 2006; McLean, Mattiske, & Balzan, 2016). However, there remain gaps in our fundamental understanding of this complex psychiatric phenomenon, limiting the efficacy of psychological treatments and interventions, and continuing to drive research across a range of disciplines.

While there are empirically supported psychological and psychopharmacological interventions aiding in the treatment of positive symptoms of schizophrenia (namely, the presence of delusions and hallucinations), not all patients are able to find effective relief from persistent and/or distressing delusions (Craig et al., 2004; Garety et al., 2011; Zimmermann et al., 2005). Treatments such as metacognitive training (MCT; Moritz & Woodward, 2007), cognitive behavioural therapy (Beck, 2011), and antipsychotic medications (for a meta-analytic review, see Haddad & Correll, 2018) have each experienced varying levels of success in alleviating positive symptoms, but 34% of sufferers remain treatment resistant (Potkin et al., 2020). Even people that report symptom reduction (i.e., in the sense of reduced distress or impairment) often persist in inappropriate levels of belief in the delusions formed during the acutely psychotic period (Kapur, 2003). In summary, for many people, delusions continue unaffected or only partially alleviated after treatment.

For people with delusions who are unable to find relief through psychological or pharmacological treatments, the overconfidence with which they hold their beliefs remains firm. Inappropriate conviction is clearly an important feature of delusions and can present a barrier to their effective treatment (Bitter et al., 2015; Moritz, & Woodward, 2007; So et al., 2013). However, while there are some mechanisms demonstrated to induce illusory or inappropriate conviction in healthy adults (discussed in the next section), little research has

examined whether these mechanisms could underlie the development and maintenance of delusional beliefs in clinical populations.

The truth effect

There is a rich and robust body of research demonstrating the effect of repetition on truth evaluation, whereby statements that have been seen or heard before are more likely to be judged as true (Hasher, Goldstein, & Toppino, 1977; Dechêne et al., 2010). In one classic experiment, Bacon (1979) read aloud a series of trivia statements, ranging across ten different general knowledge categories, to a group of students. The participants were informed that half of the statements were in fact true, and half were false, and were asked to rate the perceived truth of each statement on a 7-point Likert scale (ranging from "definitely true" to "definitely false"). After a 3-week delay, participants were presented with another series of statements to evaluate, half of which had appeared in the original list. It was found that repeated statements received significantly higher truth ratings on their second occurrence than on their first and were rated higher in truth than the newly presented items. Essentially, participants' belief in presented information was stronger following its repetition.

This phenomenon, known as the *truth effect,* has since been demonstrated across multiple experimental designs and contexts, and has been shown to be robust against variations in delay (ranging from mere minutes up to several months; e.g., Brown & Nix, 1996), modes of presentation (i.e., auditory, visual, or both; e.g., Begg & Armour, 1991), levels of processing (e.g., passive versus active engagement at phase one; Hawkins & Hoch, 1992), and presentation time per statement (Gigerenzer, 1984; for a review, see Dechêne, Stahl, Hansen, & Wanke, 2010). The effect is typically measured as a difference score between first and second presentation ratings for a repeated item (the 'within-items effect') and/or differences in ratings between repeated items and matched novel statements (the 'between-items effect'). The effect

of repetition on judgements of truth is not limited to a binary response shift (i.e., from an evaluation of falsity to one of truth) but can also be measured as an *increase in one's confidence* in a proposition relative to their initial response (e.g., from 'maybe true' to 'probably true' or 'definitely true'). Multiple studies investigating the truth effect have required participants to indicate how true or false they believe each statement is on a Likert scale, allowing researchers to measure shifts in degrees of confidence following repetition. Notably, the directionality of the effect naturally remains stable across most contexts, whereby repetition increases the likelihood and degree to which a statement will be judged as true, rather than false (although manipulations to the contextual framing of experiential fluency to artificially link dysfluency of processing with prior exposure can, under some circumstances, flip the directionality to which fluency influences subsequent judgements; see Unkelbach, 2007).

Processing fluency

Manipulating variables such as repetition or colour contrast likely influences a statement's perceived truth via its effect on *processing fluency* – that is, the subjective ease with which a stimulus is processed (Reber & Schwarz, 1999; Dechêne et al., 2010; Unkelbach, 2007; Unkelbach & Greifeneder 2013). For example, statements that have been encountered before (i.e., repeated) are more fluently processed than those that are novel (Jacoby & Dallas, 1981). Experimental findings have indicated that more fluently processed statements across a range of manipulations are typically perceived as being 'more true' than statements that are more difficult to process. These manipulations include increased colour contrast (Reber & Schwarz, 1999), the use of rhyming (McGlone & Tofighbakhsh, 2000), and semantic priming (Kelley & Lindsay, 1993). While ease of processing is a subjective experience, it can be objectively measured via reaction time – that is, more fluently processed information is processed faster than information that is disfluent (Reber, Wurtz, & Zimmerman, 2004). For

example, Reber and Schwarz (1999) manipulated fluency by presenting statements in varying colour contrasts, measuring ease of processing by reaction time. Statements in a high contrast (i.e., highly visible) colour were taken to be more fluently processed than low contrast (i.e., moderately visible) colour statements, which was supported by the observation that reaction times following exposure to the former were significantly faster than those for the latter. They concluded that high contrast statements were significantly more likely to be judged as true than those in low contrast and argued that processing fluency drove this measured bias in truth judgments.

As repeated statements are also processed more fluently than novel statements, facilitated processing fluency may also underlie the effect of repetition on truth judgments. Fluency manipulations appear to modulate a wide variety of judgments, including judgements of liking (Westerman, Lanska, & Olds, 2015), familiarity (Whittlesea, 1993), confidence (Alter, Oppenheimer, Epley, & Eyre, 2007), and source credibility (Begg et al., 1992; Henkel & Mattson, 2011). Notably, in each case the directionality is maintained, with increased fluency leading to stimuli being more favourably assessed (i.e., more likely to be true, credible, or liked), and manipulations of reduced fluency leading to stimuli being assessed more negatively (Griffiths & Mitchell, 2008).

While processing fluency persists as an important and valid explanation for the repetition-induced truth effect, there have been other accounts proposed to drive this phenomenon. Some have argued that conscious perception of repetition (even if erroneous) is necessary to elicit an increase in perceived truth (Bacon, 1979), while others have pioneered an account based on familiarity (Begg et al., 1992), or highlighted the role of convergent validity (Arkes et al., 1991), or the importance of subjective frequency (Hasher, Goldstein, & Toppino, 1977) in increasing truth judgements. Essentially, while each of these accounts share

some similarities, they can be differentiated by their predictions regarding the roles of explicit memory, conscious perception, or metacognitive feeling/sensation elicited by the experience of repetition.

Although these alternative accounts can be discriminated under precise experimental conditions, it is not clear that each account is qualitatively distinct from a fluency-based account of truth judgments when viewed more broadly across truth effect experiments as a class (see Unkelbach & Griefeneder, 2013). This is because fluency may be the underlying driver of each account, thereby placing fluency as the indirect cause of truth perception via a range of more proximal causes (such as familiarity, subjective frequency, or convergent validity). In other words, fluency is perhaps first experienced as a metacognitive sensation upon encountering a stimulus, and then interpreted as that stimulus being either familiar, recently encountered, frequently encountered, or otherwise. For example, increased fluency of processing has been shown to induce feelings of familiarity (e.g., Whittlesea, Jacoby, & Girard, 1990), and subjective frequency (e.g., Tversky & Kahneman, 1973), which, by extension, provokes a sense of convergent validity (Arkes et al., 1991). Ultimately, each of these interpretations of the feeling of fluency may then be used to inform truth judgments, thereby leading to an (indirect) effect of perceived fluency on truth assessments.

A more recent theoretical model acknowledges the nuanced contribution of explicit and implicit cues in judgements and decision-making. This new account (Unkelbach & Rom, 2017) posits that judgement is in fact informed by multiple cues, including fluency. This model – the *referential theory* – allows for the role of conscious information such as recognition, available information, and explicit knowledge, as well as unconscious metacognitive experiences such as fluency, in informing judgements and decision-making (Unkelbach & Greifeneder, 2018). The key idea underlying this is that implicit (e.g., fluency) and explicit cues (e.g., knowledge)

might jointly activate previously established referential networks in memory. For example, upon repetition, the statement, “On every continent there is a city called Rome” might trigger a coherent network of previously associated words, “continent”, “city” and “Rome”, and this match with memory is used as information regarding its veracity.

Essentially, this referential model allows for the incorporation of the role of direct knowledge and information in the effect of repetition on truth, as fluency alone cannot account for this effect where explicit information is present. It also allows for the incorporation of a multitude of factors previously implicated in the truth effect phenomenon as jointly informing truth judgements, rather than arguing the role of any one factor as the key mechanism. This theory is consistent with the fluency account and does not negate its role in the truth effect; rather, it provides an added framework as to *why* or *how* fluency might logically lead to an increase in perceived veracity. Furthermore, it has been demonstrated that experiential fluency plays a role in truth judgements even in the presence of declarative information, suggesting that explicit knowledge and fluency jointly influence judgements of truth (Unkelbach & Greifenefer, 2018). Surprisingly, Unkelbach and Greifeneder (2018) reported that fluency continued to influence truth regardless of the accuracy or trustworthiness of the accompanying information, further demonstrating the strong and robust role of processing fluency in everyday judgements and decision-making. This new model acknowledges the important role of fluency effects in truth evaluation, while also allowing for the role of explicit knowledge and memory effects.

Intuitive reasoning

Judgements can be unknowingly and involuntarily influenced by implicit or unconscious cues, including processing fluency, regardless of whether one has access to explicit knowledge (see, for example, Unkelbach & Greifeneder, 2013). In the context of the

truth effect, even if participants are aware that only half of the statements presented to them are true – while half are false – repeated statements are still more likely to be judged as true than novel ones (Dechêne et al., 2010). Such a mechanism has all the hallmarks of the fast-working, automatic system of cognition commonly referred to as *intuitive reasoning* (Kahneman, 2011). There is no logical reason why a statement that is easier to read (e.g., displayed in a high colour contrast) might be more likely true than a statement that is more difficult to read (e.g., in low colour contrast). The reasoning behind such an assumption, therefore, is ultimately not rational.

Intuitive vs rational reasoning

Cognitive and social psychologists often approach judgement and reasoning as having two key systems: intuitive and analytical (Alter, Oppenheimer, Epley, & Eyre, 2007). Appropriate use of each of these reasoning styles is highly adaptive and beneficial, with different circumstances calling for the use of different approaches in everyday judgements and decision-making (Kahneman, 2011). Intuitive reasoning is often referred to as 'Type 1' thinking, and is described as fast and effortless, subtle, independent of intention, and to occur with little awareness by the individual employing it (Topolinski & Strack, 2009). It is also thought to encompass associative and implicit learning – as a skill or piece of knowledge becomes well-known, it becomes automatic and effortless, and the use of conscious effort is no longer required (Stanovich & Toplak, 2012). This type of thinking is often utilised in situations where we have little or no direct information available (e.g., when inferring the intentions of others; Baker et al., 2009), where a task is implicitly known (e.g., when driving a car), or where there is insufficient time to make a deliberate and consciously informed judgement (e.g., responding in dangerous situations). This form of reasoning is susceptible to bias, as mental shortcuts favour efficiency at the cost of accuracy.

In contrast, analytical or rational inference is often referred to as 'Type 2' thinking, and is a slow, explicit, effortful, and deliberative process. Type 2 thinking includes the use of strategies such as hypothetical reasoning and cognitive simulation (Stanovich & Toplak, 2012), and involves the use of explicit knowledge in forming judgements and decisions. This form of inference is capable of overriding intuitive thinking when careful, precise judgements are required, when time is not a crucial factor, and when information is either available or can be readily sourced (although it must be acknowledged here that intuitive reasoning can influence people in analytical reasoning conditions, and vice versa). Therefore, one could say that rational inference favours accuracy over efficiency.

The role of processing fluency in intuitive reasoning

Topolinski and Strack (2009) argued that processing fluency informs intuitive judgements, whereby increased or unexpected fluency is used as a cue for determining, for example, whether a triad of words is semantically coherent. In one of their experiments, they investigated whether the feeling of ease in processing a triad of coherent words (e.g., SALT, DEEP, and FOAM), was used to establish whether those words shared a common fourth word relative (e.g., SEA). This was determined by measuring individuals' response latencies to reading the word triads (fluency) and observing whether those triads were accurately judged as coherent. The authors found that coherent triads were processed significantly faster than incoherent triads and argued that feelings-of-ease act as a cue in intuitive judgements of coherence.

In the context of the truth effect, where participants are unaware of the truthfulness of a statement (ambiguity), intuitive reasoning processes are employed and perceptual fluency or disfluency (as elicited by the presence or absence of repetition) becomes a proximal cue for truth (Dechêne et al., 2010). However, as previously noted, the truth effect has also been shown

to be robust against explicit information (Fazio et al., 2015), and this can occur even in situations where information is labelled 100% accurate (Unkelbach & Greifeneder, 2018). These findings indicate that intuitive or unconscious reasoning processes influence judgements together with explicit knowledge, even in situations where we might reasonably expect intuitive processes to play no role at all because they are redundant. This finding should not be taken lightly, as it as it points to the potentially significant role of intuitive reasoning processes in determining how much subjective conviction a person holds for a belief, even when there is good, external evidence available for that person to determine that the belief is true (or false), without resorting to intuition. Such a mechanism provides a possible explanation of how a person might inappropriately believe something, even in the presence of high-quality evidence contradicting that belief.

Delusions and the truth effect

Intuitive reasoning processes play an important role in assessments of truth and are influenced by subtle, subjective aspects of our experience, such as how easily we process a piece of information (Unkelbach & Greifeneder, 2013). These processes are implicit and can influence our judgements even in the face of valid cues or explicit information (Unkelbach & Greifeneder, 2018). Given this, it would seem possible that disruptions to intuitive reasoning processes could have widespread, undetectable influences on our assessments of truth – even in the presence of conflicting evidence.

Despite the body of literature surrounding the truth effect, there has been little research investigating how the phenomenon might relate to the development of clinical delusions in psychosis. There appear to be potentially important parallels. Firstly, the truth effect characteristically increases the perceived truth of a statement, and this effect persists even when people have good evidence that the target statement is true or false (Unkelbach & Greifeneder,

2018) or that their experience of repetition provides no additional support for the statement (e.g., Bacon, 1979). Similarly, the maintenance of inappropriately high levels of conviction for a belief even in the presence of contradictory information is arguably the most diagnostic feature of a clinical delusion (American Psychological Association, 2013). Finally, the growing evidence for the effectiveness of a metacognitive approach to treating delusions – metacognitive training – suggests that intuitive processes are an important feature underlying clinical delusions. Therefore, it remains an interesting question as to how intuitive processes affecting judgements of truth – such as repetition fluency effects – might play a role in the development of delusional beliefs. Given the growing evidence for the robustness of the repetition-induced truth effect in the face of explicit knowledge, it is reasonable to consider that this phenomenon might also play a role in the maintenance of such beliefs in the face of contradictory evidence.

Several studies have investigated the role of implicit cues in the development of false beliefs (e.g., Balzan et al., 2012; Grimmer et al., 2022; Pennycook & Rand, 2020), and a few studies have explicitly examined interactions between processing fluency effects and delusion formation (e.g., Corlett et al., 2009). One study in particular directly assessed whether clinical delusions were associated with a greater propensity to experience the truth effect. Moritz et al. (2012) measured the effects of repetition on truth judgements in both a normal population and those with a probable diagnosis of schizophrenia. Participants completed the Community Assessment of Psychic Experiences (CAPE-42; van Os, Verdoux, & Hanssen, 1999), a 42-item self-report measure developed to tap into three important dimensions of psychosis: positive symptoms (such as hallucinations and delusions), negative symptoms (such as avolition and social withdrawal), and depressive symptoms (such as sadness, hopelessness, and feelings of guilt) (Stefanis et al., 2002). The to-be-judged statements presented included a series of general

trivia and a series of ‘delusion-relevant’ items (for example, claims relating to government surveillance and conspiracy beliefs). Researchers observed a standard truth effect across the range of items, in both populations. That is, repeated items were rated as more likely to be true than novel statements. Furthermore, in those with a probable diagnosis of schizophrenia, there was a significantly higher truth effect observed for delusion-relevant items.

While this study provided some interesting preliminary findings, there were some methodological limitations that challenge broader interpretations. These limitations were noted by the authors at the time. For example, there were an unequal number of items in each statement condition (only 8 delusion-relevant versus 30 neutral items), these items were not explicitly balanced on features relevant to memory performance (e.g., true/false status, familiarity), they used a relatively small sample size, only a single measure of delusion proneness was used, and the delusion-relevant items almost exclusively focused on persecutory themes. These limitations – which will be discussed in more detail in Chapter 2 – together with the absence of published replications of this finding, highlight the need for further investigation.

The conceptual thinking behind the potential relationship between delusion formation and the truth effect is as follows: An anomalous experience, such as an unexpected experience of fluency, occurs at an unusual moment – for example, a person looks at you on the street and appears very familiar to you, but you do not know them. Such an experience, if salient enough, leaves a feeling of unease or uncertainty, prompting the need for a suitable explanation. This attempt to explain the occurrence might come in the form of an unusual or unlikely thought – for example, “They are following me”. If unusual experiences like this unexpected fluency (interpreted, in this case, as familiarity) continues over time, the idea might be repeated, forming the basis of a delusion. Of course, this theory does not explain the initial experience

of fluency. In the above example, it may be that this familiar person *is* in fact someone you have seen before (e.g., an employee at your local grocery store), and the feeling of familiarity is indeed accurate, but interpreted incorrectly. Alternatively, it could be that the initial anomalous experience is internally facilitated, due to unusual activity in neurotransmitters such as dopamine (see Meltzer & Stahl, 1976). In any case, the subsequent repetition of the attached idea, for those who are susceptible, might be a facilitator in the development of the delusional belief.

Current aims

The aim of the proposed research will be to understand the role of processing fluency and the truth effect as it relates to the development of conviction in delusions. (For the purposes of the present thesis, the word 'conviction' will be conceptualised as a form of attitude strength (APA, 2018); one that sits on a continuum, rather than being defined purely by the existence of the highest level of identification with an idea. The use of this word in the current APA definition of delusions prompts its use in the present research, although its conceptual relation to the word 'certainty' here is noted, and therefore may, at times, be used synonymously with certainty.) This aim can be understood as three distinct questions, which are considered in turn.

First, are statements related to the most common themes of delusions particularly susceptible to intuitive reasoning, such as the truth effect? This question was investigated in Chapters 2 and 3 using a large, unselected sample of adult participants. In Chapter 5, the related issue of personal relevance was explored. Many delusional themes involve a web of beliefs that are ultimately about the self; this is most evident in delusions of reference, in which the core attribute is the misperceived personal relevance of an event (Startup, Bucci, & Langdon, 2009). Chapter 5 explored whether the magnitude of the truth effect was influenced by whether a statement was about oneself or about a valued other. To anticipate, the present experiments

revealed the truth effect paradigm itself to be highly influenced by experimental parameters: the truth effect was strongly influenced by the choice of statements shown and the questions asked. This complicated quantification of the truth effect, so item-based analyses and parametric studies were also conducted to yield a sufficiently balanced and sensitive measure of the truth effect to reveal any effects of content type. This was explored in Chapter 3.

The second motivating question concerned whether there were important individual differences in susceptibility to the truth effect. This was explored by examining the relationship between the delusion-proneness dimension and the magnitude of experienced truth effect. This analysis pooled data across the experiments in the earlier chapters to yield a sufficiently large sample (N = 302) to offer confidence in our estimate of the relationship, if any, between delusion proneness, schizotypy and the truth effect.

The third research question was essentially the interaction of the first two questions. That is, the research examined whether those people more prone to intuitive reasoning were more likely to demonstrate that capacity for statements with delusion-relevant themes. Chapter 4 directly assessed this question using a series of hierarchical multiple regression analyses. Finally, all findings are collectively discussed in the General Discussion (Chapter 6).

Chapter 2: The Effect of Belief Content on the Truth Effect

There is a rich and robust body of research demonstrating the effect of repetition on truth evaluation, whereby statements that have been seen or heard before are more likely to be judged as true (Hasher, Goldstein, & Toppino, 1977; Dechêne et al., 2010). This phenomenon – known as the truth effect – has been demonstrated across multiple experimental designs and contexts, and has been shown to be robust against variations in delay (ranging from mere minutes up to several months; e.g., Brown & Nix, 1996), modes of presentation (i.e., auditory, visual, or both; e.g., Begg & Armour, 1991), levels of processing (e.g., passive versus active engagement at phase one; Hawkins & Hoch, 1992), and presentation time per statement (Gigerenzer, 1984; for a review, see Dechêne, Stahl, Hansen, & Wanke, 2010). However, there has been little research investigating the role of different types of *content* in the effect of repetition on truth – that is, the influence of the theme or subject matter of a given piece of information on its propensity to elicit a truth effect.

To date, most research in this area has generally focused on trivia (e.g., Hasher et al., 1977), opinion statements (e.g., Arkes et al., 1989; Riesthuis & Woods, 2024), health claims (e.g., Sundar et al., 2015), or false claims in the media (e.g., fake news, Pennycook et al., 2018; and misinformation, Vellani et al., 2023). While these findings provide a strong basis for investigating the truth effect in response to other types of information. However, there is still little understanding as to how various themes might elicit different responses in truth evaluations – that is, do some themes or ideas elicit a *greater* perception of truth when repeated (or more fluent) than other types of information? This chapter seeks to investigate this question, focusing on thematic content related to commonly held delusional beliefs, and comparing with neural trivia claims such as those pertaining to science, history, and geography.

Thematic content of delusions

A common observation when looking at the content of clinical delusions is the high prevalence of certain themes (Freeman & Garety, 2014; Kiran & Chaudhury, 2009). It remains an open question as to why certain types of beliefs tend to be overrepresented in clinical delusions. One possibility is that some types of information are processed differently than other types of information, so that distinct reasoning strategies are employed in response to different subject matters. This chapter explores whether some ideas – such as those commonly expressed in delusions – might be more susceptible to evaluation by intuitive processes than others, simply by virtue of their content.

Some types of beliefs are so prevalent in delusions that they are used as classification categories based on the presence or absence of that content (Kiran & Chaudhury, 2009). For example, persecutory delusions are commonly identified among patients with schizophrenia and delusional disorder and are characterised by a patient's underlying belief that an organisation, group or individual is out to threaten, harass or harm them (Bentall et al., 2001; Kiran & Chaudhury, 2009); grandiose delusions involve the belief that one is superior in some way – be it by inflated worth, possession of special powers, or being especially famous or important (Knowles, McCarthy-Jones, & Rowse, 2011); and delusions of reference involve the perception that seemingly unrelated or neutral external events are somehow personally significant and meaningful (Startup, Bucci, & Langdon, 2009).

It is notable that, upon first glance, all of these themes involve claims for which it is difficult to objectively derive direct evidence. These are the conditions under which intuitive reasoning might reasonably be expected to dominate (Kahneman, 2011). Persecution involves the intentions of others, and intentions are inferred. In practice, delusions of reference involve the feeling that a particular object or event was intended as a sign or omen of significance for

the patient. The significance or personal meaning of an object or event is indirectly inferred, which leads to difficulties in challenging those beliefs in cognitive therapy. Who is to say, for example, that the placement of a nearby purple vase does not have special meaning or significance? Finally, one's own abilities or powers, which are the subject of grandiose delusions, may be measured against objective yardsticks or standard measures in relatively formal settings, but informally are very often a matter of perception and inference (see for example, the Dunning-Kruger Effect; Kruger & Dunning, 1999). Perhaps the frequency with which such delusions are identified among clinical populations is due to these claims being particularly prone to intuitive appraisal, which over time may lead to the form of inappropriate conviction that defines a delusion.

Although there is currently little research investigating the interactions between processing fluency effects, statement content, and delusion formation, there is one study that jointly addresses these constructs. As discussed in Chapter 1, Moritz et al. (2012) investigated the effects of repetition on truth judgements in both nonclinical and clinical populations. Specifically, they measured the magnitude of the truth effect in 40 healthy participants and 36 individuals with a probable diagnosis of schizophrenia. The researchers hypothesised that an exaggerated truth effect would be observed for those in the schizophrenia group, potentially contributing to the formation and maintenance of psychotic symptoms such as delusions. They also examined whether there was a difference in the effect of repetition on truth for emotional (i.e., delusion-relevant) statements and neutral (i.e., factual knowledge) statements. Neutral statements included, for example: "On each continent there is a town called Rome", while emotional statements included, "The German federal police use approximately 3000 cameras for the purpose of video-based face-detection".

All participants completed a 42-item survey called the Community Assessment of Psychic Experiences (CAPE), which taps into psychosis-related phenomena including positive, negative, and depressive symptoms (van Os, Verdoux, & Hanssen, 1999). Participants were presented with a mixture of emotional and neutral statements at the initial test phase. One week later, participants were again presented with a series of emotional and neutral statements: some of which had appeared in the initial test phase, and some new statements. Results replicated the overall truth effect: statements that were repeated received higher truth ratings than new statements. Interestingly, correlational analyses revealed that patients higher in positive symptoms exhibited a significantly larger truth effect for emotional/delusion-relevant items relative to healthy participants and those low on positive symptoms. Furthermore, for participants with a probable diagnosis of schizophrenia, a significant correlation was found between scores on self-report scales for delusions and the truth effect for emotional items.

While this study provided some interesting preliminary findings, its limitations – noted by the authors themselves – prompt further investigation. First, due to the recruitment of a clinical population, it cannot be ruled out whether delusion-relevant items held personal relevance for some participants, nor whether the emotional valence of the items influenced responses (rather than their emotionality alone). For example, if some participants suffered from persecutory delusions directly relating to government surveillance, they would be much more likely to endorse those trivia items relating to government surveillance. As a result, it cannot be ascertained whether the larger truth effect for delusion-relevant items was due to the content alone, or due to pre-existing beliefs. Secondly, and relating to this issue, the simple category of 'emotional/delusion-relevant' was rather broad and not well-defined. This indicates that specific subcategories of delusional content were not addressed directly, with most items clustered around the theme of persecution. Thirdly, the unequal number of items per condition

(30 general trivia vs. 8 delusion-relevant) meant the influence of measurement error was higher in one condition (delusion-relevant) than the other (neutral trivia). Specifically, a single, outlier item could be expected to have an outsized effect in the delusion-relevant category relative to the trivia category. Equal presentation of items per statement type will ensure greater control in replication studies. Finally, this study does not address which types of statements might influence judgements of truth specifically – it merely investigates whether a difference might exist between neutral and non-neutral stimuli. This final point is the most crucial, in that this study will seek to further elucidate the characteristics of various statement types (relating to common delusional themes) and their influence on the size of the truth effect in a normal population.

The focus of this chapter will be on further investigation of the finding of an exaggerated truth effect among delusion-relevant content as reported by Moritz et al (2012). To foreshadow later points of discussion and enquiry, it is important to note that an exact replication of Moritz's study is not conducted here, in that we do not separate participants into clinical and nonclinical groups for evaluation. (The reason for this is further outlined in Chapters 4 and 6.) However, this chapter only seeks to explore the role of content – independent of individual differences – on truth; that is, the effect of thematic content on the size of the truth effect will be investigated independently of any differences in delusion-proneness here.

Nevertheless, it is possible that we may see a higher truth effect in delusion-relevant items even in a non-clinical sample, if we consider, for example, the prevalence of belief in conspiracy theories among the general population. The content of such strongly held beliefs, while not delusions in the clinical sense, share similar themes to those in delusions (in fact, one of the subcategories of delusion-relevant statements is 'conspiracy beliefs'). In addition, we know that belief in fake news has proliferated among the population – surviving regardless of

implausibility (Fazio et al., 2019). We therefore may find that people, in general, respond to the effect of fluency more greatly in the face of delusion-like content.The analysis of individual differences is withheld until Chapter 4.

Aim and hypotheses

The aim of the current investigation was to better understand the role of thematic content on the magnitude of the truth effect, using repetition as the primary manipulation of processing fluency. The experiment was conducted online and involved recruitment of participants from the general population. Specifically, this chapter investigated whether statements relating to delusional content were more susceptible to the effects of repetition on truth judgement over two experiments. The two experiments were very similar, but varied in the question participants were asked. For repeated items in Experiment 1, the participants were asked to rate their truth using an identical question on both occasions. In Experiment 2, participants rate the interestingness of statements on their first presentation and truth on the second presentation.

Across both experiments a standard truth effect was expected, whereby repeated statements should receive significantly higher truth ratings, on average, than novel and to-be-repeated statements. Second, it was predicted that a greater increase in perceived truth will be observed following repetition of statements relating to delusional thinking than for neutral items. In addition to the two broad categories of thematic content, the present experiments were designed to probe five different subcategories of each statement type. For delusion-relevant statements, subcategories included: superstitious belief, conspiracy theories, persecution/citizen surveillance, biological contamination, and complementary and alternative medicinal practices. These items were compared against neutral control statements grouped into science, history, geography, pop culture and general knowledge.

Overview of delusion-relevant subcategories

The non-obvious presumption of connections between things and events are a common feature in delusions (Dowdy & Graves, 1994; Hemsley, 1993), for example, observing a dark cloud in the sky and deducing that one will die the following day (Kiran & Chaudhury, 2009). Similarly, *superstitious thinking* involves unusual or unlikely interpretations of causality or perceptions of relatedness between seemingly unrelated things and events, such as knocking on wood to prevent something unwanted from happening (Risen, 2016; Vyse, 2013). *Conspiratorial beliefs* also involve sensationalistic interpretations of events. These beliefs tend to oppose the status quo understanding of societal issues, such as vaccination, climate change, and political agendas (Pytlik, Soll & Mell, 2020). In fact, more recently, belief in conspiracy theories has been linked to schizotypal traits, intuitive thinking style, and paranoia (Barron et al., 2018; Pytlik, Soll & Mehl, 2020).

Persecution and citizen surveillance items contain themes directly related to common persecutory delusions, such as the belief that an individual or government entity is out to observe, control and/or harm them (Bentall et al., 2001; Kiran & Chaudhury, 2009). *Biological contamination* is a less common theme among delusional beliefs, however, is a hallmark of delusions of a somatic nature such as Ekbom's syndrome, a belief that one's body has been infected or infested by parasitic organisms (Kansal, Chawla, & Singh, 2012). Finally, *complementary and alternative medicine (CAM)* refer to practices used to restore or support health that have not been scientifically proven and as such are not used in traditional medicine (Tabish, 2008). While links between delusional beliefs and the use of CAM have not been directly determined, there appears to be an association between use of CAM and propensity towards magical thinking (Kikan, 2019). Magical thinking is another phenomenon, like

superstition, that is commonly observed among both people with schizotypal traits and people with schizophrenia (Garcia-Montes et al., 2014).

Any list of delusion themes is bound to be controversial, but the goal here is not perfect coverage. Instead, the logic of the experiment merely requires that the delusion-relevant categories are *more* relevant than the control categories (e.g., geography), and sufficiently related to delusions (in at least many instances) to engage the same form of processing. Further, the intention of using defined subcategories of delusion-relevant ideas related to specific clinical delusions (and subclinical traits) was that it allows some investigation of whether particular themes may selectively engage intuitive or heuristic processing during truth evaluation.

We hypothesized that delusion-relevant statements that are more closely analogous to the content of clinical delusions should be more susceptible to the effects of repetition than less directly related statements. Specifically, delusion-relevant items should elicit stronger truth effects than control items, and statements associated with high-frequency delusion themes (e.g., persecution) will show the largest truth effects. For example, items relating to persecution might yield a larger truth effect than those regarding alternative medicine, as themes of persecution are commonly observed within clinical delusions, whereas themes of alternative health beliefs are less typical.

Experiment 1

Method

Participants

This study was approved by the Flinders University Human Research Ethics Committee (ID: 2671). Participants were recruited online via Prolific Academic online recruitment

platform (https://prolificacademic.co.uk/). Prolific has been shown to be more reliable than alternative recruitment sites, yielding higher quality data, greater diversity of participants, higher participant naïveté regarding experimental design, and fewer cases of malingering and failed attention-checks (Palan & Schitter, 2018; Peer et al., 2017).

Demographic data from one hundred and two participants (56 male) were obtained from Prolific. Mean age of respondents was 33.33 years (SD = 13.28), and the majority (92%) were based in the United Kingdom, USA, Canada, and Australia. One third of participants were students (n = 34), and all were listed as having English as a first language (as per the recruitment requirements). Participants received a payment of 3.13GBP for approximately 25 minutes of their time (7.51GBP/hour).

To determine whether this number of participants would provide sufficient power to reduce the chances of Type II error, a power analysis was conducted using G*Power (version 3.1; Faul et al., 2009). It was determined that a total sample size of 88 would be sufficient at the level of .95 (or 54 participants at the .80 benchmark previously suggested; Cohen, 1988; Cohen, 1992) for detecting an effect size of .39 (as per Dechêne et al., 2010) for within-item differences in truth ratings between novel and repeated items (i.e., the within-items truth effect). Therefore, we can conclude that the following truth effect experiment, along with all subsequent truth effect experiments in the following thesis, are sufficiently powered.

The experimental survey was created and distributed using Qualtrics survey engine (https://www.qualtrics.com), whereby participants were able to access the survey from their personal device. Qualtrics formats surveys to be easily undertaken on a computer, tablet, or mobile phone.

Materials

The Short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE).

The short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE) was developed as a self-report measure assessing schizotypal personality traits (Mason, Claridge, & Jackson, 1995). The 43-item questionnaire taps into four distinct categories representative of schizotypy: Unusual Experiences, cognitive disorganisation, introvertive anhedonia and impulsive non-conformity. While providing less detailed information regarding a person's propensity towards schizotypal personality, the short version of the O-LIFE takes less time for participants to complete and is considered a valid and reliable measure when compared with the long version and other measures of schizotypy (Fonseca-Pedrero et al., 2015).

The Community Assessment of Psychic Experiences (CAPE-42). All participants additionally completed a 42-item survey called the Community Assessment of Psychic Experiences (CAPE-42), which taps into psychosis-related phenomena including positive, negative, and depressive symptoms (van Os, Verdoux, & Hanssen, 1999). While the O-LIFE provides better indications of schizotypal personality, the CAPE provides information regarding delusion-proneness. The measure has been used widely, and has been shown to be a stable, reliable, and valid measure of subclinical psychotic symptoms (Konings et al., 2006).

Trivia statements. All appropriate neutral and delusion-relevant stimuli used in Moritz et al.'s (2012) study were translated into English and included in this experiment. Items that were removed from the original study were those that (in the author's judgment) have since become more likely to be known, e.g., "Unlike an epidemic, a pandemic is not locally limited", or that did not generalise to a non-German audience, e.g., "The oldest still existing German soccer club was founded in Berlin". To expand upon the previous experimental materials, additional statements were obtained from a broad range of internet sources. As previously

stated, additional neutral stimuli were comprised of trivia statements covering five key areas: science, history, geography, pop culture and general knowledge. Additional delusion-relevant items were developed to cover five key areas relating to thematic delusional content: superstitious belief, conspiracy theories, persecution/citizen surveillance, biological contamination, and complementary and alternative medicinal practices. While some of these themes are more typically observed within delusions (e.g., persecution and contamination), others are not as common (e.g., culturally bound superstitious beliefs). The idea was to cover alternative beliefs over a range of categories, with some more closely relating to clinical delusions than others.

All items were based on true statements that were specifically chosen for their obscurity and ambiguity, to minimize the influence of pre-existing participant knowledge on truth evaluations. Half of the obscure facts were presented directly, thereby leading to items whose correct answer was "true." The other half were altered to render them unequivocally "false". Superstitious belief items, for example, were created by collecting a series of superstitious cultural beliefs from countries around the world and adjusting to render half of them false (e.g., in the true statement, 'Sleeping with one's head to the North is considered bad luck in Japan', *North* was changed to *South*). This strategy of starting with a true item and then altering it to create a false statement was used to minimize any systematic differences in familiarity that might occur between true and false statements if, for example, the false statements were novel, manufactured statements. A similar strategy was utilised across all other categories to render half of all statements false. Neutral items were taken from trivia websites covering various topics. Delusion-relevant statements were taken from trivia pages, government websites, and scientific papers (e.g., findings regarding the effects of yoga on cortisol, chiropractic intervention on pain, and so on). A full listing of all statements can be found in Appendix A.

An informal, internal pilot and review process was conducted within our laboratory group, and this resulted in the 96 statements that appeared in the final survey. Half were neutral items and half were delusion-relevant. Orthogonally, half of the statements were true, and half were false. Approximately equal numbers of true and false statements appearing in each of the five neutral and five delusion-relevant categories (there were 9 or 10 items per category).

Design

The effect of repetition on truth ratings was measured using a within-subjects design, in which presentation of neutral and delusion-relevant statements were manipulated to appear either once or twice throughout the survey. The study had a 2 (repetition: repeated, not repeated) x 2 (statement type: neutral, delusion-relevant) repeated measures design. The dependent variable here was the truth rating response given for each statement, with the truth effect for repeated items calculated by subtracting the truth rating given for the second presentation of a statement from the truth rating from non-repeated items presented interleaved with the repeated items. Measures of individual differences were also captured by the schizotypy/delusion-proneness questionnaires. Results from these measures were calculated for each participant and separated into different subcategories of schizotypy, and correlational analyses were used to measure whether these results predicted truth ratings for delusion-relevant or neutral items, or the size of truth effect itself.

Procedure

Participants were informed that the study aimed to examine how briefly presented words affect how people think, to better understand how and why people form rapid, intuitive judgements about the truth of a statement.

After consenting to participate in the experiment, participants were asked to disclose their age and gender. Participants then completed the O-LIFE and the CAPE-42, which consisted of a series of statements to which participants would respond by indicating whether they experienced a particular phenomenon, and/or how often this occurred. They were then provided with instructions informing them that they would be asked to evaluate the truth of a series of trivia questions, and to indicate whether they believed each question was true or false using the response options provided. They were asked not to consult other sources or to spend too long on each statement, but to use their intuition to decide what they thought the answer might be.

Participants completed the entire truth ratings questionnaire in one sitting, with one statement presented per page, along with a Likert scale below each statement. A meta-analysis has shown the truth effect to be independent of the length of delay between presentations, with an interval of mere minutes sufficient to produce an effect (Dechêne et al., 2010). Therefore, in the present experiment, there was no delay between presentation of the first and second series of statements. In the first half of the experiment, participants rated 64 novel statements, as per the study by Moritz et al. (2012). A 6-point scale has been shown to produce a larger truth effect, as it removes the option for a neutral response (Dechêne et al., 2010). Directly following the initial set, another 64 statements were presented: half of which had been encountered in the first half, and half that were novel. The order of presentation was randomized within each set (i.e., the initial set was randomized independently of the second set of items). Overall, one third of all items were repeated, totalling 32 statements (half neutral, half delusion-relevant; and of each category, half true, and half false), so that each participant rated the truth of a series of 128 statements in total. Three different counterbalancing conditions were created to ensure each of the items was equally likely to be assigned each of the three

possible roles: novel item in the initial set, novel item in the secondary set, repeated item (and thus shown in both sets).

The truth effect can be calculated in two quite distinct ways, depending upon the control used in the relevant comparison. The present experiment was designed to make both comparisons possible within a single study. Specifically, the truth effect can be determined in a "within-item" manner by comparing truth ratings after the first presentation of a statement with ratings given for that same statement after repetition. The other way is the "between-item" effect, whereby truth ratings for statements that have been repeated are compared against truth ratings for novel statements within the same set of items.

Calculation of the within-items truth effect was obtained by subtracting response ratings for to-be-repeated items from ratings for repeated items. Calculation of the between-items truth effect was obtained by subtracting truth ratings for repeated items from ratings for novel items in the second half of the survey.

Results

Reaction times

We removed all responses taking more than 10 seconds, along with their corresponding paired response for repeated items (i.e., for repeated statements with missing response data at first or second presentation, the corresponding item response was also removed). This was to control for participants who may have consulted other sources (e.g., Wikipedia) in responding to the trivia questions, or disengaged from the task generally. In addition, responses that took less than one second were removed to control for inattention. Overall, 10.55% of all truth rating responses were removed from the final dataset, leaving a total of 11,818 item responses for analysis. To check that these data-cleansing processes did not affect the conclusions drawn, all

analyses reported below were repeated for the full data set (with no item removals). No differences in the pattern of significant findings were observed.

Delusion-proneness measures

The Short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE). O-LIFE scores were determined by adding up affirmative responses (or negative for items with reverse coding) over four sub-scales: Unusual Experiences, Cognitive Disorganisation, Introvertive Anhedonia, and Impulsive Nonconformity. Population comparisons were made with norms published by Mason et al. (2005) using z-scores. A z-score is a standardised score which transforms data so that comparisons between two population means can occur (Crawford & Howell, 1998). Z-scores are representative of the number of standard deviations from which one dataset differs to another, and significant p-values indicate a significant difference between population samples as reflected by this z-score.

For Unusual Experiences, the average score was 4.01 (SD = 2.91), with a z-score of .22, p = .41. The mean score for cognitive disorganisation was 5.45 (SD = 3.46), z-score = .36, p = .36. The means for introvertive anhedonia and impulsive nonconformity were 3.49 (SD = 2.60), z-score = .51, p = .30 and 3.28 (SD = 2.23), z-score = .34, p = .37, respectively. In summary, the sampled population did not differ markedly from the normative population, suggesting that there was nothing remarkable about the sample.

The Community Assessment of Psychic Experiences (CAPE-42). Table 1 depicts means and standard deviations for each of the three dimensions. The highest average score lay on the positive dimension, which is most related to positive symptoms of psychosis (including delusions and hallucinations). Z-scores were computed to compare our population means with the healthy population CAPE-42 score averages reported by Moritz et al. (2012). This revealed that the weighted frequency scores from each population are broadly comparable (z-scores:

1.06, p = .15, -.03, p = .51 and -.02, p = .51 for positive, negative, and depressive symptom frequency, respectively).

Table 1

Means (and Standard Deviations) for scores on the CAPE by Frequency and Distress, and by Positive, Negative, and Depressive symptoms.

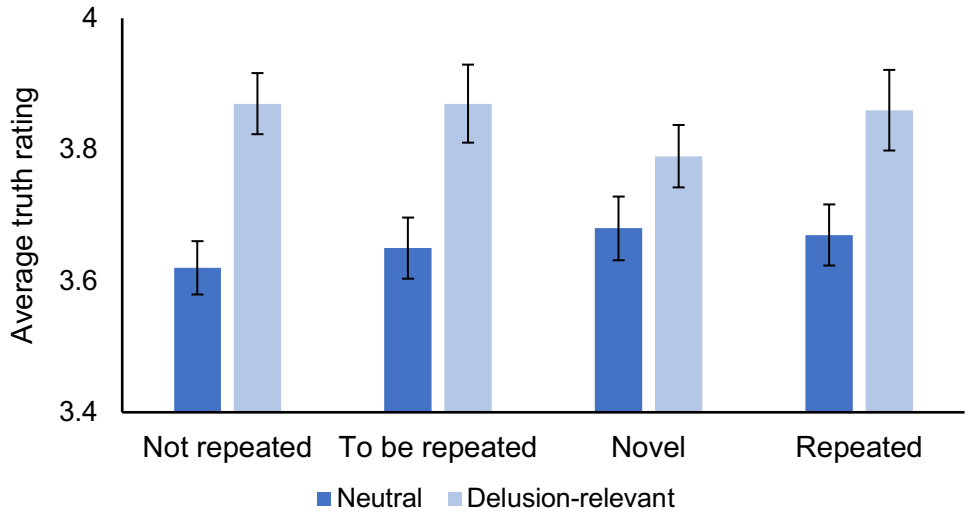
	Positive	Negative	Depressive
Frequency	1.40 (.30)	1.73 (.44)	1.77 (.40)
Distress	1.68 (.60)	1.85 (.66)	2.08 (.79)

The truth effect

Truth effect is calculated by subtracting truth ratings for repeated statements from ratings for control, novel statements. However, there are two valid ways to calculate this score (Dechêne et al, 2010). The first analyses examined differences in ratings between the first and second presentation of the same item (a "within-items" effect). That is, comparing the ratings for an item when it is repeated against ratings for that same item when it was (previously) novel. Secondly, and more commonly, you can compare ratings for novel and repeated statements shown concurrently in the second half of the experiment (a "between-items" effect).

Figure 1

Average truth ratings (with standard deviations) for neutral and delusion-relevant items across training and test phases.



Between-items effect. Differences in average truth ratings between novel and repeated items in each statement condition are depicted in Figure 1. To analyse these data, we used a repeated measures ANOVA with two factors: statement type (delusion relevant or neutral) and repetition status (repeated or novel). Contrary to our predictions, no significant differences between novel and repeated statements were observed (irrespective of statement type), $F(1,101) = .91$, $p = .34$, $\eta^2 = .002$. There was a significant main effect of statement type, as delusions relevant items were rated as more truthful than neutral statements, $F(1,101) = 8.06$, $p = .01$, $\eta^2 = .04$). There was no significant interaction between statement type and repetition, $F(1,101) = 1.29$, $p = .26$, $\eta^2 = .003$.

Within-items effect. An analysis was carried out to investigate whether a difference in truth ratings occurred between first and second presentation of the same item (i.e., within-items truth effect). A repeated measures ANOVA was used, with two factors: statement type (delusion relevant or neutral) and repetition status (repeated or to be repeated). The main effect

of repetition on truth ratings from first to second presentation was not significant, contrary to expectations, $F(1,101) = .13$, $p = .72$, $\eta^2 < .001$. This means that there was no significant change in truth ratings from the first to second occurrence of the same statement – i.e., no within-items truth effect was observed. The main effect of statement type was significant, showing a difference in overall truth ratings between delusion-relevant and neutral items, $F(1,101) = 9.42$, $p = .003$, $\eta^2 = .07$. The interaction between statement type and repetition was not significant, $F(1,101) = 1.58$, $p = .21$, $\eta^2 < .001$. In summary, no evidence of a truth effect was observed in this experiment, either in the between-item comparison or in the within-item comparison.

Exploratory analyses

The following analyses were conducted in an attempt to gain a greater understand of the null results observed in this experiment. See discussion section for further justification and explanation of these analyses.

Consistency in responses between first and second presentation of repeated items (within-subjects) was investigated in an item-by-item analysis. The analysis revealed that 55% of repeated statements received identical truth ratings from first to second presentation. Upon removal of these consistently rated items (i.e., leaving only those item responses that differed between first and second presentation), another t-test was conducted, to investigate whether any truth effect occurred for inconsistently rated items. This analysis revealed that there was no significant difference between first ($M = 3.75$, $SD = .67$) and second ($M = 3.71$, $SD = .60$) presentation of these remaining items, $t(95) = .65$, $p = .52$, $d = .07$. Indeed, this was further corroborated by the observation that, among inconsistent statement responses, the proportion of repeated statements with decreasing truth ratings was equal to that of repeated statements with increasing truth ratings (49.7% and 50.3%, respectively).

Discussion

The truth effect is generally reported as robust and has been replicated repeatedly over decades and across multiple experiments by different research teams. Nevertheless, we failed to find a truth effect in Experiment 1. Several post hoc explanations were considered. At first it appeared possible that the delay between first and second presentation of a given statement may have been too brief. However, the specific delay used by other within-items studies is not routinely reported, so it was difficult to precisely determine whether this was the cause of the present null effects. More specifically, delay magnitude is typically reported in the literature when the initial presentation and the repeated presentation occur on different days or in different sessions (e.g., Bacon, 1979; Brown & Nix, 1996), but it is less often reported at the specificity of minutes when both items were presented within a single session (e.g., Unkelbach, 2007). Moreover, the most comprehensive review of the truth effect literature (Dechêne et al, 2010) found that truth effects could be robustly observed in a single session, which led to the present experimental design.

Fortunately, we recorded stimulus onsets for every item and were thus able to determine between-item delays for every repeated item, for every participant. Across all trials, there was an average delay of less than 9 minutes between first and second presentation of repeated items (M = 8.96, SD = 3.77), and 85% of trials had a delay of less than 13 minutes between first and second presentation of a given statement. Of the studies that explicitly reported delay between study and test, one study (Nadarevic & Erdfelder, 2014) found no truth effect between items separated by 10 minutes but found robust effects when they were separated by a week. In contrast, a second recent study reported a series of significant small-to-medium truth effect sizes with an average delay of 10.75 – 16.75 minutes between item presentations (Unkelbach

& Greifeneder, 2018). So, while it remained possible that the delay was too brief to elicit an effect, the present delay period was also comparable with that used in other, successful studies.

However, in their study of the effect of item delay on truth effect magnitude, Nadarevic & Erdfelder (2014) found that a second factor importantly moderated any impact of delay. Specifically, they noted that when participants were asked to make the same rating at the first and the second presentation of a statement (as in the present study), the delay between first and second item presentation became crucial. However, if they were asked to make different ratings at the first and second item presentation, then delay had a much smaller impact upon the observed truth effect magnitude, and that a truth effect could be observed with short (10 minute) delays. Nadarevic & Erdfelder (2014) proposed that if people were asked to make the same judgment about the same stimuli in rapid succession, that they would attempt to be as consistent as possible and that this drive for consistency would override the influence of intuitive biases, such as the truth effect. Indeed, they observed a much higher (61%) consistency rate in truth ratings across statements in the ten-minute delay condition of the experiment than when longer delays were used.

There is some evidence to support the operation of this kind of process in the present data. Participants were asked at the conclusion of the experiment what they thought the reason for statement repetition might be. Of these, more than half (54.9%) guessed some type of cognitive mechanism, with most responses relating to memory or the experimenter checking for consistency. Specifically, 32.35% of participants suggested the reason may have been to measure people's consistency across ratings, while another 9.8% believed this was to assess memory. Surprisingly, 12.75% guessed some influence of repetition on truth. For most participants, these assumptions regarding cognitive mechanisms may have led participants to aim for consistent responses, while the short delay allowed for relative ease of retrieval of

previous responses. Consequently, it is reasonable to suggest that this desire for consistency in responses across the experiment contributed to the null effect observed.

To further investigate this hypothesis, an item-by-item analysis of truth ratings for repeated statements was conducted (see Exploratory analyses in Results section above). as noted, the analysis revealed that 55% of repeated statements received identical truth ratings from first to second encounter, providing some evidence for memory and consistency effects. Furthermore, the proportion of repeated statements with decreasing truth ratings was more or less equal to that of repeated statements with increasing truth ratings. When consistent responses were eliminated, no significant difference in truth ratings between first and second presentation of statements was observed. In other words, even for items that were not rated consistently from first to second occurrence, there was no truth effect.

This finding can be interpreted in two ways: Firstly, one could argue that the lack of a truth effect for inconsistently rated items suggests that consistency in fact did *not* drive the null results observed. As inconsistent responses indicate an absence of memory and consistency effects, they should theoretically show an effect of repetition on truth, which did not occur here. Alternatively, even though almost half of responses were not consistent, one could argue that the lack of a significant difference in ratings for these items could reflect an *attempt* to remain consistent coupled with a degree of noise in the response process. That is, participants may feel that a statement is true with 40% conviction, but due to random fluctuations in their internal state, sometimes they give that item a rating of 3, sometimes 4 and sometimes 5. When accurate recall of initial ratings was absent, but the underlying conviction remains, they might try to recall what they would have rated given that level of internal conviction, again choosing a value somewhat randomly between 3-5. Under such circumstances, extreme values would tend to regress to the mean, leaving approximately half of the inconsistently rated items going

up and half going down, exactly as was observed. Experiment 2 was designed explicitly to minimize the influence of the drive for consistency across phases.

Experiment 2

The rationale for Experiment 2 was to make some methodological changes to the original experimental design to remove the opportunity to respond consistently and thereby ensure the generation of a truth effect. One way to do this is to increase the delay between sessions. This should reduce people's memory for their initial responses, and instead force them to rely on intuition or evaluative processes. However, introducing a long delay or multi-session task would constitute a significant departure from Experiment 1. Further, it would reduce our control over participants' pre-existing knowledge, as they may research some or all of the stimulus statements in the period between sessions.

Fortunately, Fazio and Sherry (2020) have detailed an alternative approach that reduces the need for imposed delays. Fazio and Sherry (2020) replaced the initial truth judgement from their task with a prompt for participants to indicate how *interesting* they found each statement. Then, in the second half of the experiment, participants were asked to rate the truth of each statement (half of which had been previously encountered in the ratings of interest task) and compared truth ratings for repeated statements against novel, control statements. The cost of this approach is that it removes the ability to quantify the truth effect using the within-items method, but the benefit is that the truth effect (quantified using the between-items method) appears to large and robust (Fazio & Sherry, 2020; Fazio, Rand, & Pennycook, 2019). This approach is further supported by reports that merely thinking about truth at time 1 can reduce the magnitude of the truth effect (Newman et al., 2020). Therefore, the present experiment adopted this approach.

Method

Participants

This study was approved by the Flinders University Human Research Ethics Committee (ID: 2671). One hundred new participants (53 female) were recruited through Prolific. Mean age of respondents was 32.68 years (*SD* = 10.68), and the majority (88%) were based in the United Kingdom, USA, Canada, New Zealand, and Australia. One quarter of participants were students (*n* = 25), and all were listed as having English as a first language. Participants received a payment of 3.13GBP for approximately 25 minutes of their time (7.51GBP/hour).

Procedure

Measures, stimuli, and design are identical to Experiment 1. Two changes were made. First, all statements were evaluated for interestingness in the first half of the task (on a scale from "Not interesting at all" to "Extremely interesting"). Second, evaluation of interestingness at first presentation to encourage attention to statements while avoiding the consistency in responses seen in Experiment 1. Three questions were imposed between the first and second half of the experiment. Given that participants were asked to make different judgments in the first and second half of the task, some degree of pause and explanation was required. In this period, participants were also asked three questions related to the perception of time spent doing the experiment up until that point. These questions were not important for the experiment but aided by creating a short break before the second set of questions and to help obscure the purpose of the study. No other changes were made to the original method, with all counterbalancing measures and statements remaining the same as Experiment 1. Only subjects who reported English as their first language were recruited, and participants received a payment of 3.75GBP for approximately 30 minutes of their time (7.51GBP/hour).

Results

Reaction times

The same data-cleaning approach was used for Experiment 2, with participant responses to statement evaluations that took more than 10 seconds or less than 1 second removed before the final analysis. Overall, 10.48% of all truth rating responses were removed from the final dataset, leaving a total of 11,459 item responses for analysis. Any participant with more than 20% of data removed were to be excluded, however no participant met this criterion. Once again, removal of this data did not alter the pattern of significance observed.

Delusion-proneness measures

The Short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE). Scores over most categories were slightly above previously documented population norms (Mason, Linney, & Claridge, 2005; Fonseca-Pedrero et al., 2015). For Unusual Experiences, the average score was 5.14 (SD = 3.28), with a z-score of 0.61, p = .27. The mean score for cognitive disorganisation was 6.12 (SD = 2.99), z-score = .59, p = .28. The means for introvertive anhedonia and impulsive nonconformity were 3.87 (SD = 2.27), z-score = .70, p = .24, and 3.34 (SD = 2.29), z-score = .37, p = .36, respectively.

The Community Assessment of Psychic Experiences (CAPE-42). For the CAPE, the corresponding frequency response for missing distress response data fields were removed. Table 1 depicts means and standard deviations for each of the three dimensions. The average weighted score for frequency of positive symptoms was 1.58 (SD = .43), while the weighted average for negative symptom frequency was 2.06 (SD = .55), and the average for depressive symptom frequency was 2.06 (SD = .60). Z-scores were again computed to compare our population means with the healthy population CAPE-42 score averages reported by Moritz et

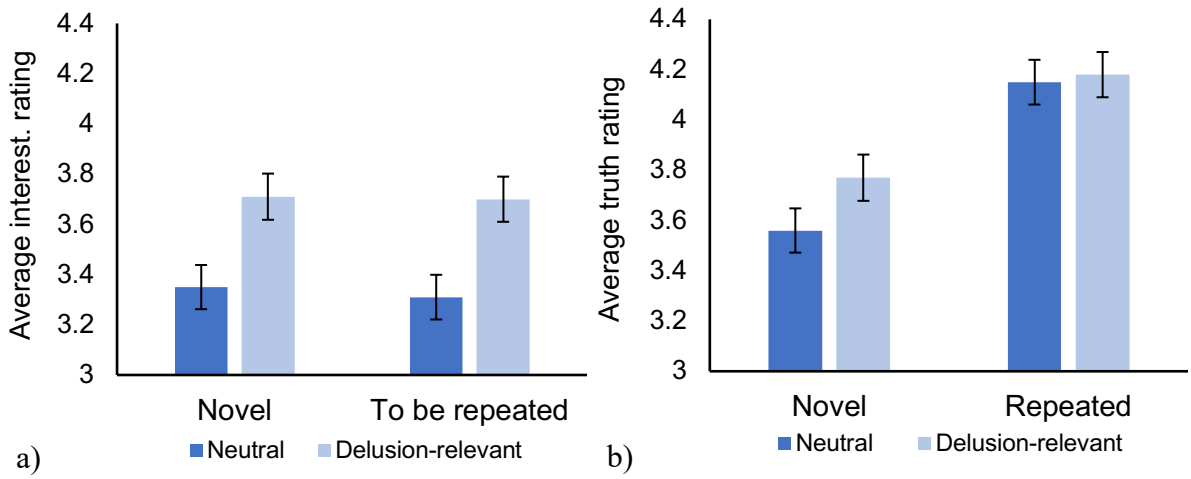
al. (2012). This revealed that the weighted frequency scores from each population are comparable for negative and depressive symptom frequency (*z*-scores: .82, $p = .21$, and .61, $p = .27$, respectively). Interestingly, our sample had an average Positive Symptom Frequency of over 2 standard deviations above the Moritz et al. (2012) population scores (z-score = 2.05, $p = .02$). Discussion of this observation is withheld for Chapters 4 and 6.

The truth effect

As previously noted, only a between-items truth effect was measurable given the design changes implemented in Experiment 2. This means that truth ratings for repeated items were compared against truth ratings for novel items.

Figure 2

Means (and standard deviations) for ratings for combined, neutral, and delusion-relevant items across (a) training (i.e., 'interestingness' evaluation) and (b) test (i.e., 'truth' evaluation) phases.



For a depiction of average truth ratings for novel and repeated statements across statement types, see Figure 2. As can be seen, there exists a marked difference in average truth

ratings between novel and repeated items, signifying a truth effect. There also appears to be almost identical mean truth ratings across delusion-relevant and neutral items once repeated, in contrast to non-repeated conditions. A 2 (repetition: repeated, not repeated) x 2 (statement type: neutral, delusion-relevant) repeated measures ANOVA was carried out to investigate whether a difference in truth ratings occurred between novel and repeated statements in the second half of the experiment (i.e., between-items truth effect). Consistent with our initial predictions, the main effect of repetition on truth ratings was significant, producing a large truth effect, $F(1,101) = 9.54$, $p < .001$, $\eta^2 = .22$. In addition, the main effect of statement type was significant, showing a difference in overall truth ratings between delusion-relevant and neutral items, $F(1,101) = 9.54$, $p = .003$, $\eta^2 = .01$. The interaction between statement type and repetition was also significant, showing a small effect, $F(1,101) = 7.65$, $p = .007$, $\eta^2 = .007$.

Post-hoc t-tests were conducted to take a closer look at the initial ANOVA findings. A t-test compared the magnitude of the truth effect (repeated minus novel items) in delusion versus neutral items, and found a larger effect in the neutral items, $t(101) = 2.77$, $p = .007$, $d = .28$. Interestingly, delusion-relevant items elicited significantly higher truth ratings than neutral items when novel, $t(101) = 4.17$, $p < .001$, $d = .29$; whereas there was in fact no difference in truth ratings between neutral and delusion-relevant items after repetition, $t(101) = .69$, $p = 1.0$, $d = .07$. Contrary to our hypothesis, repetition exerted a smaller influence on delusion relevant than neutral items, possibly due to the higher novel ratings for delusion relevant than neutral items.

Accuracy

Given the main effect of truth ratings for delusion relevant and neutral items, it was possible that people had differential pre-existing knowledge of the two stimulus categories. For example, perhaps they knew that more of the (actually) true conspiracy statements than the

(actually) true geography statements. Such a pattern could appear as a smaller truth effect for delusion-relevant items on this between-items metric, similar to that observed.

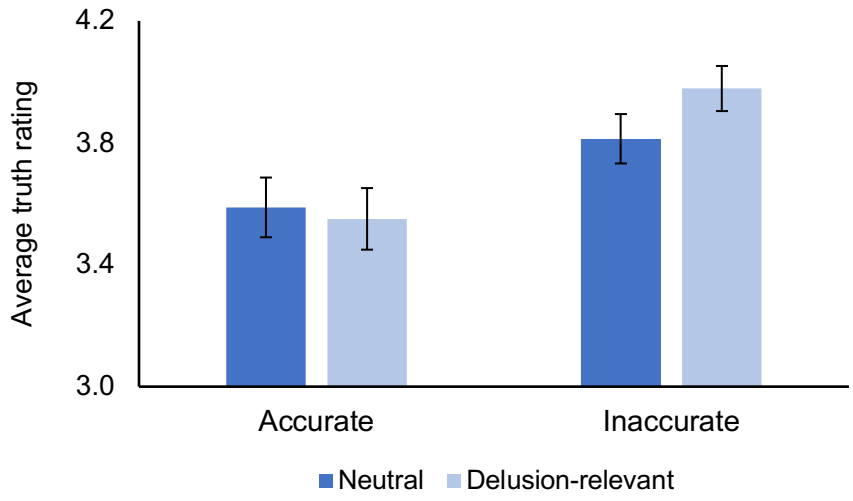
To control for any effects of differential levels of accurate knowledge between stimulus categories on the size of the truth effect, an additional ANOVA was conducted using *only* inaccurate responses. To anticipate, this analysis revelated the same pattern of findings. A 2 (statement type: delusion-relevant, neutral) x 2 (repetition status: repeated, novel) revealed a significant main effect of statement type, $F(1,98) = 9.08$, $p = .003$, $\eta^2 = .01$, and repetition status, $F(1,98) = 35.50$, $p < .001$, $\eta^2 = 0.09$. Again, the interaction was also significant, $F(1,98) = 9.32$, $p = .003$, $\eta^2 = 0.01$.

Further post hoc comparisons revealed there was no significant difference in truth ratings between neutral and delusion-relevant truth ratings for accurate responses, $t(1,99) = .61$, $p = .54$, $d = .06$, but a significant effect of statement type remained for inaccurate responses, $t(1,99) = 3.11$, $p = .004$, $d = .31$.

In summary, when people knew the correct response (or at least guessed correctly), no difference in truth ratings between statement categories was observed. However, amongst their inaccurate responses we saw the same pattern in the overall data, which included a significant effect of statement type. This pattern argues strongly against an influence of differential prior knowledge explaining the observed main effects and interactions but does suggest that when people did not know the veracity of a delusion-relevant statement, they were more likely to guess “true” than following control statements.

Figure 3

Mean truth ratings for accurate versus inaccurate responses in Experiment 2.



The effect of delusion-relevant item category on truth

Within the delusion-relevant item type, there were 5 different categories: Persecution, Contamination, Conspiracy, Supernatural Belief, and Alternative Medicine. Analyses explored the pattern of truth evaluation within each category, to see whether there were significant differences between them. Arguably, some of these categories (such as Persecution) were more related to common clinical delusions (e.g., delusions of persecution), than others (e.g., alternative medicine). We hypothesised that items relating to Persecution would elicit larger truth ratings and a larger truth effect than other categories in the delusion-relevant condition. We also predicted that scores in positive schizotypy subscales would correlate positively with truth ratings for Persecution delusion-relevant statements.

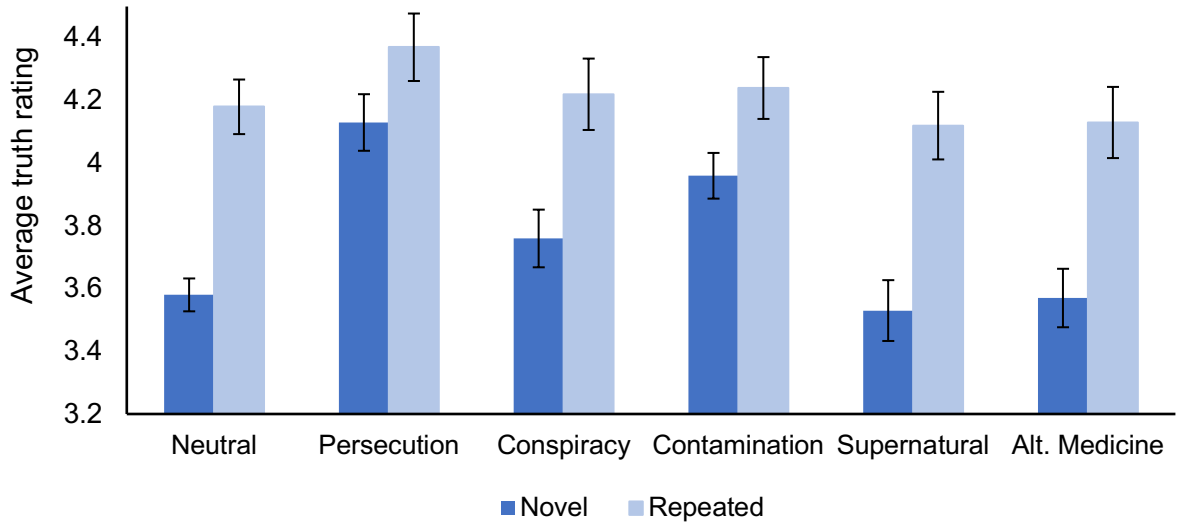
A repeated measures ANOVA was conducted with two factors: item category (Persecution, Contamination, Conspiracy, Alternative Medicine, Supernatural Beliefs, and Neutral), and repetition status (repeated, novel). This revealed a significant main effect of item category, $F(5,430) = 8.64$, $p < .001$, $\eta^2 = .04$ (see Figure 4). Again, a difference between

repeated and novel statements was significant and clearly observable, $F(1,86) = 31.78$, $p < .001$, $\eta^2 = .08$. The interaction between item category and repetition was also significant, revealing a small effect, $F(5,430) = 2.54$, $p = .03$, $\eta^2 = .01$.

To explore this interaction, each of the individual delusion-relevant sub-categories were compared against the average of the five neutral categories with respect to the magnitude of the truth effect (the difference in truth ratings between repeated and novel items). This resulted in five additional simple interaction effects. Post hoc t-tests with Bonferroni corrections revealed that the truth effect in the Persecution and Contamination items was significantly smaller in magnitude than that seen in the Neutral items, $t(430) = 2.52$, $p = .01$, $d = .33$ and $t(430) = 2.25$, $p = .03$, $d = .31$, respectively. In contrast, the truth effect seen in the other three delusion-relevant categories did not significantly differ from that seen in the Neutral category: Supernatural Beliefs, $t(430) = .04$, $p = .97$, $d = .01$, Conspiracy, $t(430) = 1.01$, $p = .31$, $d = .14$, and Alternative Medicine, $t(430) = .25$, $p = .80$, $d = .03$. As can be seen in Figure 4, the observed difference in truth effect magnitude for the persecution and contamination items stems primarily from the initial ratings given to the delusion relevant statements, rather than from any particular facilitation of the repetition effect for these categories.

Figure 4

Average truth ratings (with standard error bars) comparing novel and repeated items across Delusion-Relevant item categories and all Neutral items.



Interestingness and truth

Correlations were conducted to test whether there was a relationship between ratings of interestingness in the first half of the experiment and ratings of truth in the second half. Specifically, we compared those items that were presented in both phases of the experiment (and therefore, repeated), to see whether initial ratings of interestingness were related to subsequent ratings of truth. No relationship was observed between Interestingness and Truth for repeated Delusion-Relevant items (r = .11, p = .29), or for Neutral items (r = -.05, p = .63). This indicates that interestingness ratings were not predictive of evaluations of truth for subsequently repeated items.

Moritz et al. (2012) items and truth

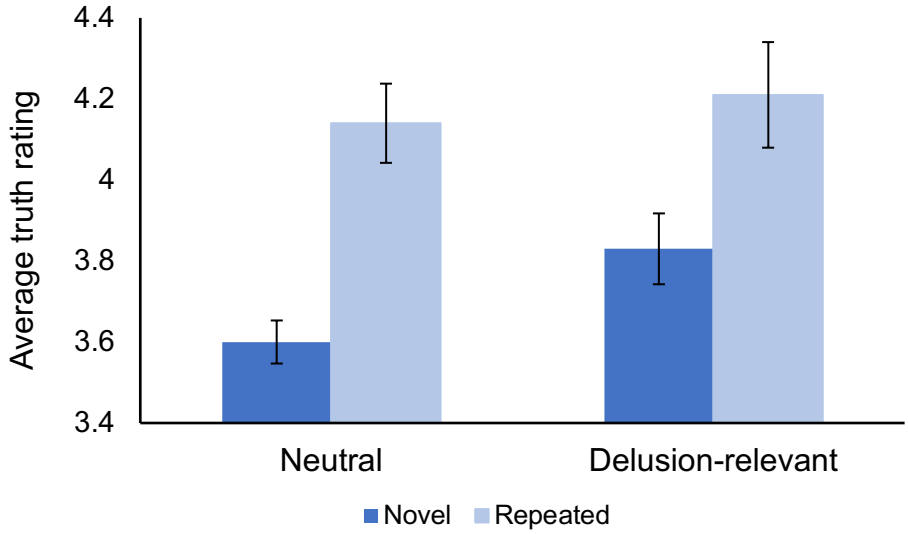
To offer a more precise replication of the truth effect study published by Moritz et al. (2012), analyses of the truth effect in Experiment 2 were conducted using only those items taken from the original study. Items utilised from Moritz's (2012) study included 24 Neutral

and 7 Delusion-Relevant items (from a possible pool of 55 Neutral and 16 Delusion-Relevant statements).

As can be seen in Figure 5, the difference in truth ratings between novel and repeated items suggests a truth effect. This pattern of results resembles those seen in Figure 4, where all items were included in the analysis. Including truth ratings only for those items used by Moritz et al., a 2 (repetition: repeated, not repeated) x 2 (statement type: neutral, delusion-relevant) repeated measures ANOVA was carried out. The main effect of repetition on truth ratings was significant, indicating a large truth effect overall, $F(1,33) = 21.41$, $p < .001$, $\eta^2 = .20$. The main effect of statement type was not significant, suggesting there was no overall difference in truth ratings between delusion-relevant and neutral items, $F(1,33) = 2.62$, $p = .12$, $\eta^2 = .02$. This runs counter to initial findings by Moritz et al., who reported a significant difference in ratings between item conditions. The interaction between statement type and repetition was not significant, $F(1,33) = 1.08$, $p = .31$, $\eta^2 = .006$. Limiting the present data to only those items used in Moritz et al (2012) did not result in a replication of Moritz et al.'s (2012) key findings. The difference in findings does not appear to be attributable to difference in the stimuli used.

Figure 5

Means (with standard error bars) showing differences in truth ratings between novel and repeated neutral and delusion-relevant items taken from Moritz et al. (2012).



Discussion

The adjustments made to our original experimental procedure appear to have been successful, with the addition of a brief mid-experimental task and shift from truth to interestingness evaluation in the first experimental phase leading to unambiguous detection of a truth effect.

The truth effect

A large and robust between-items truth effect was detected. It was not possible to measure the within items effect because no individual item was rated on truth twice. The reduced opportunity for participants to maintain consistency in their responding (relative to Experiment 1) appeared to allow participants to rely on more intuitive responses, thereby yielding a truth effect.

Overall, delusion-relevant items were rated more true than neutral statements. However, a closer look revealed that this difference occurred only between *novel* delusion-

relevant and neutral items – there was no difference in ratings for repeated items (see Figure 2). This means that, contrary to our predictions, the truth effect was larger for neutral than delusion-relevant items.

The truth effect and statement type. It was hypothesised that items relating more closely to content seen in clinical delusions would yield a larger truth effect. In fact, Experiment 2 found a larger effect of repetition on truth for items that were neutral in content. This was likely due to a tendency for respondents to rate delusion-relevant statements as more true than neutral statements when novel. With less room for an increase in truth ratings, delusion-relevant statement endorsement may have yielded what is known as a 'ceiling effect' (Vogt, 2005). Upon closer examination of Figure 3, the pattern of results over different item categories appears to support this notion. Although there was significant variation in the ratings given to neutral and delusion-relevant statements when they were novel, there is remarkably little variation in the truth ratings given to these items after repetition. Participants' ratings for all categories appeared to converge on values between 4 and 4.3 on the scale, perhaps representing the maximal value most were willing to give to an item that they did not explicitly know the answer to. In contrast, the mean difference in ratings for novel delusion-relevant and neutral statements show a medium effect size (d = .45), indicating a non-trivial difference in baseline ratings. Together, these patterns account for the differences in truth effect magnitude seen between stimulus categories.

Another possible reason for the smaller truth effect observed in delusion-relevant items is the influence of mood. Koch and Forgas (2012) studied the impact of inducing either a negative or a positive mood on people's responses in a truth effect experiment. First, participants were exposed to a film designed to induce either a negative mood, a positive mood, or a neutral mood, before rating the truth of a series of statements that differed in fluency by

colour contrast. Their results demonstrated that inducing a negative mood can eliminate the truth effect, and the authors argued that this occurred by decreasing one's reliance on processing fluency to inform judgements. While we are without data regarding the valence of each statement in the present study, inspection of the delusion-relevant items suggests that they tended to be more negative in content than the neutral items (for example, "During Prohibition in the United States, the government literally poisoned alcohol" is a delusion-relevant item, while "The goldfish sees the widest colour spectrum of all living creatures" is a neutral item). Furthermore, the plausibility or implausibility of a claim does not appear to affect the truth effect, which is important to note regarding delusion-relevant statements (Fazio et al., 2019). If delusion-relevant items were of a more negative valence, then they may have induced a more negative mood and reduced the resulting truth effect. Note that mood manipulations of this kind are not ordinarily thought to impact one trial and not the next trial occurring seconds later, and instead tend to be manipulated for larger periods of time (minutes to hours). Nevertheless, differential mood at the time of judgment may explain the reduction in size of the truth effect seen for delusion-relevant items in Experiment 2.

A third possible explanation of the differential truth effect seen between delusion relevant and neutral items references statement length. For example, statements with a greater number of words might be interpreted as being more detailed and hence more credible at first encounter. Indeed, while sentence length has not been investigated within the truth effect literature, it has been demonstrated that mere quantity of information does affect decision-making (Keller, & Staelin, 1987). In the present study, delusion-relevant statements contained an average of 18.10 words, whereas neutral statements contained an average of only 9.83. It is possible that this difference influenced truth unexpectedly.

Finally, it is possible that participants had encountered some of the information from the statements prior to this study, which produced an increase in endorsement for those items. For example, more participants may have heard the superstitious belief regarding one's sleeping position in Japan, or the possible benefits of reiki, than about the geographical details of Ecuador or how long a python can live without food. Regardless of whether one initially believed claims about the healing benefits of reiki, if one had encountered that claim prior to the experiment it is more likely they would endorse it as true (i.e., a truth effect may have already occurred) (Fazio, Rand, & Pennycook, 2019). The post hoc accuracy data support this explanation. Participants demonstrated no differences between stimulus categories for statements that they knew (or guessed correctly), and instead all stimulus differences were localized among items that were not known. Amongst those items, delusion-relevant statements were more likely to be rated as true. That is, when a person was not able to draw on pre-existing knowledge, and instead needed to use other sources (such as the feeling of familiarity), then they rated delusion-relevant statements as more true than novel controls. This suggests a differential level of baseline familiarity between the two stimulus categories.

Diagnosing the source of baseline truth rating differences between delusion-relevant and neutral statement conditions is the primary goal of Chapter 3. The intention was to identify any stimulus imbalances between the two categories that impacted on truth ratings, and thereby develop a more sensitive measure of the influence of delusion relevance on intuitive processing.

Delusion-relevance and item category. Item categories more closely related to delusional content (e.g., persecution) were expected to demonstrate a larger truth effect than other categories (e.g., complementary and alternative medicine). In practice, the present experiment tended to find the opposite trend. However, this trend was primarily evident in

ratings for novel items, rather than in repeated items, which suggests the locus of the difference between categories was in their content, rather than the interaction of repetition with that content. Nevertheless, this observation is inconsistent with the hypothesis that certain delusion-relevant item categories would yield a larger truth effect than neutral items.

Most importantly, it must be restated here that these analyses were exploratory and lacked satisfactory power; further, the unequal number of items in each subcategory rendered conclusions from these comparisons speculative at best. For these reasons, such analyses will be largely avoided for the remainder of this thesis.

Moritz et al. (2012) items and truth

Of the original 55 Neutral and 16 Delusion-Relevant statements utilised by Moritz et al. (2012) in their preliminary study of the truth effect and the presence of symptoms and traits relating to schizophrenia, 24 Neutral and 7 Delusion-Relevant items were retained in the present study. To test whether the original findings reported were replicated in the present experiment, analyses were conducted using these 31 items only. To our knowledge, this is the most direct replication of Moritz et al.'s (2012) study to date. As illustrated in Figure 5, a significant truth effect was observed, but delusion-relevance did not exert an impact on ratings; neither the interaction nor the main effect was significant. Thus, the core findings of Moritz et al. (2012) were not replicated. Note, however, that beyond the incomplete coverage of their stimulus set, the present study had other distinct design features from Moritz et al (2012). For example, the delay period in their study was one week. As we observed in our results from Experiment 1, delay does seem to play a role in the effect of fluency on truth evaluation. Although the procedural changes in Experiment 2 led to the observation of a standard truth effect, it might be the case that a longer timeframe between repetition of items would exert a significantly different effect on truth outcomes; that is, perhaps length of delay exerts more of

an influence on the truth effect than we currently give it credit for. A comprehensive comparison between the studies and their findings is reserved for the final chapter, when data from the entire experimental series is available.

General discussion

The findings of the present chapter provide novel insight into the truth effect phenomenon and how it relates to different types of content. Although individual differences data were also collected across these experiments, the influence of delusion-proneness and schizotypy on the magnitude of the truth effect will be explored and discussed in detail in Chapter 4. In Experiment 1, we did not find a truth effect, presumably due to memory and consistency effects. In Experiment 2, altering the question across presentations led to the observation of a significant truth effect. In neither experiment did we see a difference in the expected direction, whereby delusion-relevant statements led to a stronger truth effect. In fact, in Experiment 2, the reverse effect was observed, driven largely by higher initial ratings for delusion relevant than neutral statements.

The truth effect

A key question raised by the current findings is why a truth effect was not captured in Experiment 1. A 2010 meta-analysis concluded that the truth effect was robust against multiple variations in experimental design (Dechêne et al., 2010), including the introduction of delay between item presentations and the specific question asked of participants during a to-be-repeated items initial presentation. However, these factors were more recently found to interact (Nadarevic & Erdfelder, 2014) such that they can cumulatively ameliorate the truth effect under some conditions. The authors suggested this result may have been a result of memory and consistency effects, meaning that participants appeared to remember their initial truth ratings

and replicate their responses when items were repeated. This might be further understood within a revised 'fluency-attribution' hypothesis, whereby people who can recognise and attribute experiential fluency to familiarity – for example, due to the short delay between repetitions of statements – will not interpret said fluency as a cue for truth, and thereby not increase their ratings of veracity.

This account is consistent with two other aspects of the data. First, the fact that over 40% of respondents guessed that the main purpose of repetition was a test of memory or consistency provides further support for this explanation. Second, the removal of the evaluation of truth in the first half of Experiment 1 resulted in the observation of a large truth effect in Experiment 2 (and in Nadarevic & Erdfelder's (2014) second experiment). For this reason, all subsequent experiments use an interestingness question during the first stimulus presentation.

Between-items truth effect

Using an interestingness question in the first half of the experiment limits our quantification options for the truth effect. Specifically, it means that we can only quantify the truth effect using a between-items measure. Fortunately, the between-items effect is arguably more relevant for the questions posed in this chapter. While the truth effect is almost certainly influenced by changes in levels of ease of processing (Nahon et al., 2021), difference in truth ratings *between items* is argued to be *more* dependent on processing fluency than the within-items effect (Dechêne et al., 2010). This is because the *discrepancy* between expected and actual levels of processing ease are considered as being more informative of judgements than the *absolute* level of processing ease itself (Wanke & Hansen, 2015). That is, expecting something to be fluent and finding instead that it is disfluent is more informative than the level of fluency experienced on its own. Presenting a series of novel and repeated items together in

a randomly interleaved manner maximises the opportunity for participants to detect the relative ease of the processing of the repeated items.

The effect of statement type on truth

It was hypothesised that items relating more closely to content seen in clinical delusions would yield a larger truth effect. Ultimately, the findings of the present study, both from Experiment 1 and Experiment 2, do not sufficiently answer our question regarding statement content and the truth effect. In order to convincingly answer this question, both neutral and delusion-relevant statements should start at the same level of baseline endorsement. If the two categories are not rated at the same level on initial exposure, any differential change from that baseline value between the conditions is difficult to interpret. Greater change in one category over the other here could be due to greater susceptibility to the truth effect, or it could be due to alternate variables influencing truth ratings for one or both categories. For example, suppose that people are likely to believe longer statements, emotive content, or content that feels familiar. Any content initially perceived as truthful might therefore yield smaller increases in truth with repetition. If baseline truth ratings differ based on factors such as these, and are thereby not equated, such explanations cannot be excluded. To achieve this goal, an item level analysis of the data was performed in Chapter 3.

Chapter 3: Construct Validity

Clinical delusions often contain common themes (Freeman & Garety, 2014; Kiran & Chaudhury, 2009). Beliefs containing themes of persecution or supernatural entities, for example, are frequent among clinical populations experiencing delusions, such as schizophrenia. In the last chapter, we sought to investigate whether statements thematically consistent with commonly held delusions might be more susceptible to intuitive reasoning processes. This was examined by comparing the magnitude of the repetition truth effect for delusion-relevant and neutral (i.e., non-delusion-relevant) trivia statements. It was found that, while participants rated delusion-relevant items as 'more true', on average, than novel statements, there was no difference in the size of the truth effect between statement conditions, suggesting no unique influence of repetition on delusion-relevant content. However, the presence of a baseline difference in truth ratings between novel delusion-relevant and neutral statements elicits the question as to which factors drove initial truth judgements (particularly for delusion-relevant statements). If the statements were judged differently at phase one, we cannot conclude whether or how repetition might affect judgements at phase two. Therefore, a conclusive answer to our initial research question, with respect to the effect of schizotypy and sentence content on truth effect magnitude, eluded us in Chapter 2.

This chapter seeks to explore how and why delusion-relevant items elicited higher truth ratings when they were first encountered, and thus novel, in Experiments 1 and 2. Significant effort was made to balance the statements used in Chapter 2. The statements started with a set used in previous published research (Moritz et al, 2012), which were then screened for regional specificity, and additional statements were added to balance frequency between categories. These new statements were added and classified according to sub-categories to both improve representativeness (construct validity) and structure the statement list such that between-

category content comparisons could meaningfully be made (for example, between persecutory themes and contamination concerns). Finally, an informal in-lab pilot was conducted to further balance and screen the items before deployment. Nevertheless, novel ratings between delusion relevant and neutral items were not attained.

Chapter 2 identified three variables that may have influenced initial truth ratings and thereby confounded measurement. These were: familiarity, emotionality, and statement length. The potential role of each in shaping truth ratings is considered in turn.

Familiarity

Studies have indicated that the repetition truth effect is largest between the first and second encounter of a claim, but that further exposures do have the capacity to elicit additional increases in truth ratings (Dechêne et al., 2010; Hasher et al., 1977; Hassan & Barber, 2021). Familiarity is described as a subjective feeling of recognition and has been directly associated with subjective evaluations of liking (Madison & Schiölde, 2017), affect (Garcia-Marques et al., 2016), confidence (Fitzsimmons, 2020), value (Alter & Oppenheimer, 2008), and truth (Arkes et al., 1989). While many consider familiarity as a proxy for previous knowledge, it has been argued that the *feeling of familiarity* itself – as distinct from explicit recollection of a piece of information – can drive subjective truth judgements independently of explicit knowledge and memory (Arkes et al., 1989). Indeed, it is this feeling state that is argued by some to be the primary source of the repetition effect (Begg et al., 1992). Regardless of the positions held by various experimenters in the field of truth judgements, both explicit memory *and* the feeling of familiarity have been independently shown to increase judgements of veracity among the general population (Begg et al., 1992; for a review, see Unkelbach et al., 2019).

We suspect that the delusion-relevant statements may have been perceived as more familiar than the neutral statements in Experiments 1 and 2, and that this in turn led to higher

ratings of truth for delusion relevant statements. Given the cross-sectional design of our experiment, it would be quite difficult to ascertain the relative contributions of familiarity as distinct from explicit knowledge in each of our participants, for each statement. Fortunately, it will not be necessary to tease apart the feeling of familiarity from conscious recollection or prior knowledge of a statement for the present investigation. This is because we seek to discover how 'well-known' the items in our previous truth effect study are in the general population, to test whether relative familiarity drove the observed differences in truth ratings between novel items. For pragmatic reasons, we remain agnostic about the true source of that familiarity (e.g., an encounter with a specific claim previously, explicit knowledge from a trusted source, familiarity with parts of or concepts included within a statement). The exploration of familiarity as conceptually distinct from explicit knowledge with regards to truth judgements is a complex conversation and is outside the scope of this thesis (but see Begg et al., 1992; Unkelbach & Stahl, 2009; Unkelbach et al., 2019; Unkelbach & Rom, 2017). If novel items were familiar to participants, then they may have been judged as more true. If delusion-relevant items are more familiar, on average, than neutral items, then this may account for the baseline differences in truth ratings between the two statement conditions and potentially the differential receptivity to the influence of repetition.

Emotionality

The dual-process model of judgement and decision-making largely attributes emotionality to intuitive thinking, while neutrality is associated with explicit, logical, and deliberative thinking (Kahneman, 2003; Zander et al., 2017). The content of statements elicit emotional responses to differing degree and valence (Beck & Beck, 2011), and there is reason to believe that this *emotionality* may then shape truth ratings directly (Li et al, 2022), or via differential engagement of decision heuristics (Kahneman, 2003). On reflection, the delusion-

relevant items appeared to contain semantic content that elicited a greater emotional response than neutral items, thereby facilitating a greater reliance on intuitive processes, and potentially provoking a bias for truth before repetition even occurred. Alternatively, the emotional *valence* could be more influential than to the non-directional magnitude of the affective response, whereby messages with negative and positive valences might influence truth differently. Both possibilities are valid and are supported by some promising – albeit preliminary – research (Blanchette & Richards, 2010).

Emotional arousal

While the current understanding of emotionality and its role in reasoning has not yet been extended to the truth effect per se, there has been an increasing amount of research in recent years on the relationship between affect, persuasion, and belief, particularly with regards to the spreading of (mis)information over the internet. For example, the impact of emotionality on belief in misinformation surrounding the COVID-19 pandemic (Li, Chen, & Rao, 2022), the spreading of 'fake news' online (Bago et al., 2022), and increasing proliferation of conspiracy theories (Tomljenovic et al., 2020) have all become areas of interest in psychological research in recent times.

Li and colleagues (2022), for example, recently found that increasing levels of emotionality predicted likelihood to believe in – and consequently to continue to spread – misinformation about the coronavirus pandemic online. In this study, COVID-19 risk was measured by a 7-day average of daily new deaths and new cases. People faced with a higher level of COVID-19 risk experienced both weaker positive and stronger negative emotions, and heightened emotionality (regardless of emotional valence) was correlated with increased belief in and greater likelihood to share information about the pandemic, irrespective of veracity. This

finding was of particular interest here, as the authors demonstrated that it was specifically emotional *arousal* – but not *valence* – that predicted susceptibility to misinformation.

A study by Zander et al. (2017) also found evidence to suggest that emotional arousal – but not valence – is related to the use of intuitive versus deliberative reasoning processes. In this study, the authors collected psychophysiological data by measuring participants' electrodermal activity (EDA) and responses to an affective priming procedure while using either intuitive or deliberate decision-making strategies to solve a coherence judgement task. Participants were instructed to make decisions either "quickly and spontaneously... [relying on] gut feeling... without expending any effort" (intuitive strategy) or by "thinking carefully... taking time before deciding" (deliberate strategy), when judging whether a word triad was semantically coherent (e.g., DEEP, FOAM, SEA) or incoherent (e.g., DREAM, BALL, BOOK). They concluded that the level of emotional arousal differed significantly between intuitive and deliberative judgement strategies, while emotional valence did not differ between the two strategies. While the findings of these two studies outlined do not reflect emotional arousal with regards to informational content *specifically*, they do point to emotionality as an important factor affecting judgements and decision-making.

Research examining the direct influence of emotional and neutral *content* on reasoning and decision-making processes (Blanchette & Richards, 2004; Blanchette, 2006; Blanchette & Richards, 2010) typically investigates the effect of emotional content (both positive and negative valence) and neutral content on accuracy in decision-making. In this work, participants completed a conditional reasoning task ('if *p*, then *q*') with emotional (positive, negative) and neutral contents. Participants' likelihood of providing correct responses in these reasoning tasks was compared across both emotional arousal and valence of content in a within-subjects design. The general conclusion of this work is that emotional content, but not valence,

influences effective reasoning. This has been demonstrated with statements containing emotional words (such as danger, hurt, and love), and by manipulating the emotional connotations of neutral content using positive and negative priming (Blanchette & Richards, 2004). This research is valuable as it looks at how emotional content itself – rather than merely its valence – directly influences effective decision-making and logical reasoning.

In sum, there is empirical evidence to suggest that emotional arousal can affect decision making of the same broad class as the truth judgements in the experiments of the previous chapter. Therefore, this is an important variable to consider in our attempt to understand the increased endorsement of truth among delusion-relevant items in Experiments 1 and 2.

Emotional valence

There is a wide literature spanning multiple disciplines that addresses the 'negativity bias' phenomenon in human psychology (for a review, see Rozin & Royzman, 2001). This phenomenon refers to the tendency for negative stimuli (such as events, information, social interactions, and emotions) to be more salient, potent, and therefore influential in our behaviour and cognition than positive stimuli (Baumeister et al., 2001). However, the question remains as to how this might occur; that is, does negative information affect our reasoning by eliciting more careful, elaborative thinking, or by prompting the use of more intuitive reasoning processes?

Negative valence with regards to both content and mood has been demonstrated to influence evaluations of truth (Hilbig, 2009; Koch & Forgas, 2012). Some research suggests that this is because negative stimuli are more likely to be met with greater and deeper processing, and more deliberate, careful evaluation than positive stimuli (Bohner et al., 1988; Lewicka, 1997). Subsequently, this phenomenon is thought to lead to an increase in persuasion

(Petty & Briñol, 2008). Alternatively, some argue that negative valence elicits the faster, more fluent processing and retrieval more indicative of intuitive reasoning (e.g., Taylor, 1991).

Fortunately, there is some preliminary research exploring this very idea, and how it relates to truth evaluation specifically. Recent research has investigated the differential influence of negative and positive framing on evaluations of truth, and the processes that underlie these influences (Hilbig, 2009; Hilbig, 2011). Over several experiments, Hilbig (2009; 2011) demonstrated that negatively framed trivia statements were judged as being more valid than their positively framed equivalents. For example, negatively framed statements such as, '20% of marriages are divorced within the first 10 years' was consistently more likely to receive higher truth ratings than the essentially equivalent, but positively framed statement, '80% of marriages last 10 years or longer'. Both statements convey the same information, and yet the former elicited greater judgements of truth.

The author explored the metacognitive processes that underpin this phenomenon over three follow-up experiments (Hilbig, 2011). They posited that the higher truth ratings observed for negatively framed items may have been due to either (a) a tendency for negative information to be more deeply and elaboratively processed, and this elaborate processing leads to higher persuasion of its veracity; or (b) negative information allows for faster (or more fluent) retrieval or generation of relevant and confirming evidence, thus influencing a higher perception of truth. Using response latency as the dependent variable, the author reasoned that support for hypothesis A would be reflected by a longer response time in evaluating negative statements than positive statements, because greater elaboration takes more time, and hypothesis B would be supported by a shorter response latency for negative statements than positive statements, as ease-of-processing is closely related to, and often measured by, response time (Alter & Oppenheimer, 2009). Findings accumulated over three carefully designed experiments

revealed that response latencies for negatively framed statements were significantly shorter than those of positively framed statements. These findings supported the notion that negatively framed statements allowed for more fluent processing of relevant information, and the author argued that processing fluency was the likely candidate for discrepancies in truth judgement between positively and negatively framed stimuli.

Statement length

Finally, could some unintended factor – such as statement length, which differed between statement conditions – have led participants to believe some statements to be more true than others? The average number of words per statement for delusion-relevant items (M = 18.10, SD = 6.16) was almost double that for neutral items (M = 9.83, SD = 2.96). To our knowledge, there is no previous research investigating the influence of statement length on the truth effect to date, however, we can consider pre-existing research regarding information quantity and its influence on decision-making effectiveness (Keller & Staelin, 1987), and judgements of message credibility (Kwaśniewicz et al., 2021). Recent neuroscientific studies investigating message credibility evaluation have shown that statements that are longer and more complex are significantly more likely to be judged as credible than short, simple messages. Statistically different neural activation between message types was observed in language-relevant regions, such as the angular gyrus and temporal cortex (more specifically, they observed differential activity in Brodmann areas BA.25, BA.31, BA.37, BA.39, and BA.41; Kwaśniewicz et al., 2020; Kwaśniewicz et al., 2021). Furthermore, literature on the detection of lying and Reality Monitoring Theory (Johnson, & Raye, 1981) posit that *richness in detail* (as indicated by complexity and quantity of information provided) is a signature of truth and truth-telling, and this theory is often exploited by liars in their attempts to persuade others (Nahari, 2018; Nahari, & Nisin, 2019). Due to the large difference in statement length

– and therefore, presumably, complexity – between the two conditions, it is reasonable to hypothesise that statement length (as measured by the number of words per statement) may have positively influenced initial truth ratings.

Aims and hypotheses

Firstly, the present chapter aims to elucidate the key variables driving initial truth ratings in Experiments 1 and 2 (Chapter 2), such that an understanding of why delusion-relevant items received higher endorsements of truth than neutral items is reached. This aim will be the rationale for Experiment 3. Secondly, the information obtained in Experiment 3 will inform the compilation of a new list of items that will result in balanced baseline ratings for neutral and delusion-relevant items. Experiment 4 will test these new items in a new truth effect experiment, and further investigate the role of statement content (delusion-relevant vs neutral) in the magnitude of the truth effect. We anticipate the following observations:

(a) There will be a significant positive relationship observed between familiarity ratings in the present survey, and novel truth ratings from Experiments 1 and 2; or,

(b) There will be a significant relationship observed between both the degree of emotionality (i.e., arousal) and the direction of emotionality (i.e., valence) and novel truth ratings, with higher emotional arousal predicting higher truth ratings (i.e., a positive correlation), and/or more negative valence predicting higher truth judgements than positive valence (i.e., a negative correlation); or,

(c) There will be a positive relationship between truth ratings and the number of words per statement.

Experiment 3

The following experiment seeks to elucidate the key factors driving the unequal truth ratings between neutral and delusion-relevant items in the previous chapter (Chapter 2). This involves a detailed analysis of each statement presented in Experiments 1 and 2.

Method

Participants

This study was approved by the Flinders University Human Research Ethics Committee (ID: 2671). Forty participants (24 female) were recruited online via Prolific Academic online recruitment platform (https://prolificacademic.co.uk/). The experimental survey was created and distributed using Qualtrics survey engine (https://www.qualtrics.com). Mean age of respondents was 35.54 years (SD = 13.87), and the majority (83%) were based in the United Kingdom, USA, Canada, New Zealand, and Australia. Of those with student status information listed, 20% (n = 8) were students. Only subjects who reported English as their first language were recruited, and participants received a payment of 1.88GBP for approximately 15 minutes of their time (7.52GBP/hour).

Materials

All statements used in the previous experiments were included in the present survey. This included a total of ninety-six statements, of which forty-eight were Delusion-Relevant (vs Neutral), and forty-eight were True (vs False). See Appendix A for the full list of items.

Design

This study used linear regression analyses to determine the effect of several continuous predictor variables on our dependent variable, Novel Truth Ratings. The predictor variables

were statement Familiarity, Words (calculated as a number per statement), Arousal, Valence, and Accuracy (calculated as a percentage, and used a proxy for identifying potentially 'known' answers). The variable of Truth Status (i.e., 'True', versus 'False') was also included in analyses where categorical variables were possible.

Key variables

To investigate the variables driving the baseline differences in truth ratings between Neutral and Delusion-Relevant items, all novel truth ratings recorded over Experiments 1 and 2 were pooled together. These ratings were then averaged across items (as opposed to participants), to perform an item-by-item analysis in which each statement had its own average truth rating. Note that this transition unavoidably reduced our inferential sensitivity by changing our degrees of freedom. Specifically, entering statements rather than participants meant that our inferential statistics now involved a maximum of 95 degrees of freedom (as there were 96 statements) rather than 201 degrees of freedom (as there were 202 participants), thereby reducing power. However, by pooling across experiments we could obtain more reliable estimates of the mean truth rating per sentence, thereby substantially reducing the influence of inter-individual variability and increasing the power of our item-based analyses and inferences.

Statement length was determined for each statement individually. Familiarity and emotionality ratings were obtained by asking an independent group of participants to rate each statement in the study based on its familiarity and its pleasantness on a 7-point Likert scale. The participants were new people drawn from the same global pool of participants as Experiments 1 and 2, and so would likely be representative of the same sample. Using a Likert scale enabled us to observe the degree to which participants felt they have seen an item before, and to which they found the item pleasant or unpleasant. In addition, responses to the

pleasantness scale will enable us to measure both the direction of emotionality (valence) and its extremity (emotional arousal), which will be entered as two separate variables in our analyses.

Emotional arousal and valence were calculated by first recoding the Likert scale as -3 to +3, with 0 being the "Neutral" centre rating. Emotional arousal was then determined as the average distance from 0 for any one item (ignoring the negative or positive value). For example, a score of 0 for an item would mean it had no emotional arousal, while a score closer to 3 would be considered highly emotional. For emotional valence, the negative and positive polarity was directly measured, so that items with an average score below zero denoted negative emotional valence, and those with a score above zero denoted pleasant or positively valenced claims.

Procedure

Participants were informed that the present study aimed to examine how briefly presented words affect how people think, to better understand how and why people form rapid, intuitive judgements about the truth of a statement.

After consenting to participate in the study, participants were asked to provide their age and gender. They were then instructed to read a series of trivia statements, one per screen, and to separately rate each item on a 7-point Likert scale by:

1) how *familiar* (or unfamiliar) the statement was to them, and
2) how *pleasant* (or unpleasant) they found the content of the statement (regardless of whether the statement was true or false).

Participants completed the questionnaire in one sitting, with one statement presented per page, along with two Likert scales below each statement (one for Familiarity ratings and one for Pleasantness ratings). Participants could respond to the two scales in any order, and their responses were untimed. The screen position of the scales was not randomized or

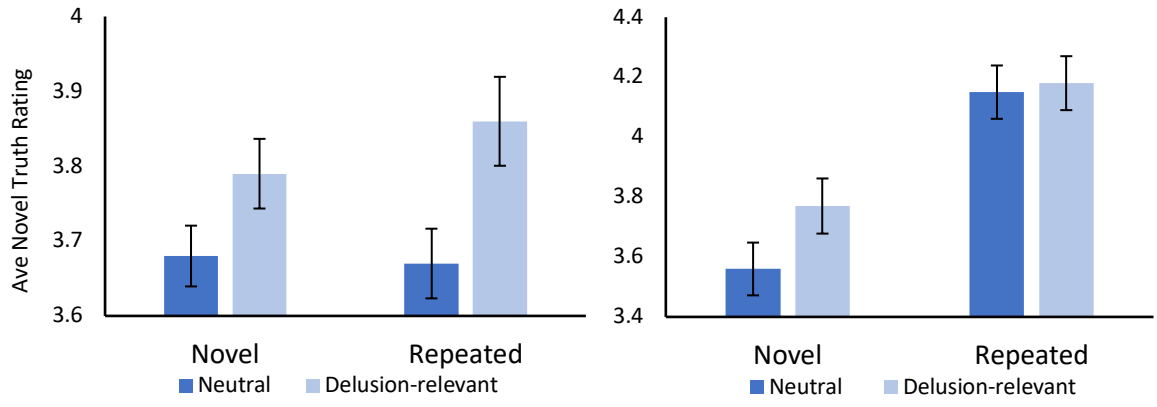
counterbalanced (familiarity ratings were always in the upper location). The order of presentation of each statement was randomised between participants.

Results

As discussed in the previous chapter, there was a significant difference observed between truth ratings for Neutral and Delusion-Relevant statements when Novel (that is, at baseline). This result was identified and discussed in Experiment 2; however, the pattern was similarly observed in Experiment 1 (see the left-hand columns in both panels of Figure 6). The baseline differences between Neutral and Delusion-Relevant items in each experiment were confirmed in two separate paired samples t-tests, indicating a medium-sized, significant difference between conditions: $t(101) = 5.16$, $p < .001$, $d = 0.51$ (Experiment 1); and $t(99) = 4.51$, $p < .001$, $d = 0.45$ (Experiment 2).

Figure 6

Average Truth Ratings (with Standard Error Means) for Neutral and Delusion-Relevant statements when Novel and Repeated (taken from Chapter 2). Left: Experiment 1. Right: Experiment 2.



To investigate the variables driving the differences between items (statements), all novel truth ratings recorded over Experiments 1 and 2 were pooled and averaged across items,

so that each statement was allocated an averaged truth rating. These variables included statement Familiarity, Words (calculated as a number per statement), Arousal, Valence, and Accuracy (calculated as a percentage, and used a proxy for identifying potentially 'known' answers).

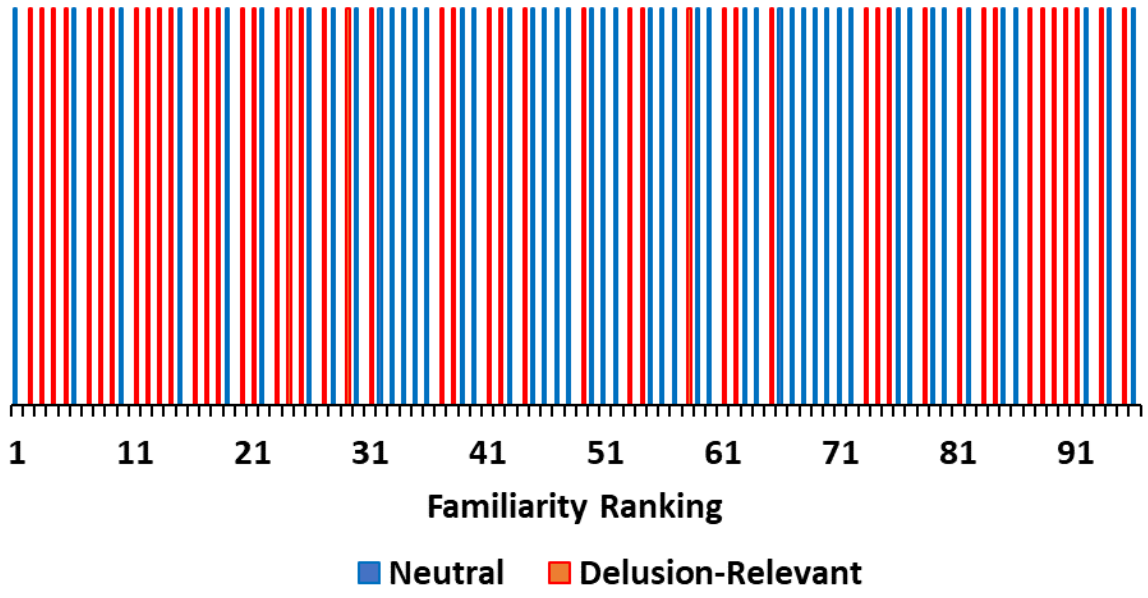
T-tests

Independent samples t-tests were conducted to further understand the differences between key variables and Statement Type (i.e., Delusion-Relevant versus Neutral). Descriptive and inferential statistics are presented in Table 1.

Figure 7 illustrates the rankings of familiarity among Delusion-Relevant and Neutral items (shown in red and blue, respectively). Looking at the figure, we can see there is a greater proportion of Delusion-Relevant items (depicted in red) on the left-hand side, suggesting several Delusion-Relevant items were rated among the most familiar. In essence, it appears that the familiarity ratings for delusion-relevant items were more extreme – that is, either highly familiar or highly unfamiliar – in general. To see whether this apparent difference in familiarity ratings between Delusion-Relevant and Neutral items was significant, a t-test was performed. Surprisingly, although there was a tendency for Delusion-Relevant items to be rated more familiar than Neutral items, the t-test determined these differences were not significant.

Figure 7

Average Familiarity ratings per item, ranked from most to least familiar. Delusion-Relevant items are depicted in red.



An independent samples t-test revealed a large, significant difference in Arousal between Neutral and Delusion-Relevant items, with Delusion-Relevant items receiving higher ratings of emotional arousal than Neutral items (see Table 1). Similarly, Valence differed between statement conditions, with a t-test revealing a large, significant difference whereby Delusion-Relevant items were perceived as more negative than Neutral items. Finally, and as expected, there was a large, significant difference in the number of words in each statement condition.

Table 2

Descriptive and inferential statistics for regressor variables (with effect size Cohen's d)

Variable	*N*	*M*	*SD*	*t*	*p*	*Cohen's d*
Novel Rating	96	3.75	.57	1.78	.08	.37
Neutral	48	3.64	.52			
Delusion-Relevant	48	3.85	.59			
Valence	96	-.47	1.08	4.50	*<.001	.92
Neutral	48	-.02	.88			
Delusion-Relevant	48	-.92	1.08			
Arousal	96	.95	.69	4.08	*<.001	.83
Neutral	48	.68	.55			
Delusion-Relevant	48	1.22	.72			
Familiarity	96	2.53	.87	1.78	.08	.36
Neutral	48	2.37	.75			
Delusion-Relevant	48	2.68	.96			
Accuracy	96	52.23	21.65	1.45	.15	.30
Neutral	48	49.05	21.13			
Delusion-Relevant	48	55.41	21.91			
Words	96	13.97	6.36	8.39	*<.001	1.71
Neutral	48	9.83	2.96			
Delusion-Relevant	48	18.10	6.16			

Note. * = significant at the $p < .05$ level, uncorrected

Correlations

An initial, uncorrected bivariate correlation analysis was conducted to investigate the strength and direction of direct relationships between each pair of variables. The results are presented in Table 3. A significant positive correlation ($r = .61, p < .001$) was observed between Novel Truth Ratings and Familiarity, showing a medium effect size. A small but significant positive relationship was also observed between Novel Ratings and Accuracy ($r = .25, p = .01$).

There were no significant relationships observed between Novel Ratings and either Arousal or Valence. Finally, the number of Words in each statement shared a small yet significant positive correlation with Novel Ratings, indicating that statements with a greater number of words tended to elicit higher truth ratings. Of note, Delusion-Relevant items had, on average, almost double the number of words than Neutral items (see Table 2 for descriptive and inferential statistics).

Table 3

Correlation coefficients for key variables

Variable	2.	3.	4.	5.	6.
1. Novel Rating	-.03	.18	*.61	*.25	*.23
2. Valence		*-.61	.17	-.14	*-.25
3. Arousal			.02	*.24	*.30
4. Familiarity				*.29	.17
5. Accuracy					*.23
6. Words					

Note. * = significant at the $p < .05$ level, uncorrected

Multiple regressions

To further investigate the role of all predictor variables in driving novel truth ratings, a series of simultaneous multiple regressions were performed. Using simultaneous regression allowed us to include one more potential predictor variable in our analyses, namely, Truth Status (being whether the statement was true or false), in addition to Familiarity, Arousal, Accuracy, and Words. Three analyses were conducted. Each analysis sought to predict Truth Ratings using the same set of variables, regressing over individual statements (rather than over participants). However, the three analyses different with respect to which statements were entered into the analysis. The first analysis looked at all statements (Delusion Relevant and

Neutral), and then two subsequent analyses looked at the Delusion-Relevant and Neutral items in isolation. This approach maximizes sensitivity to any differences in the predictive relationships within statement types, that might be obscured by pooling all statement types together. Unstandardized coefficients (beta weights) were calculated for each analysis (using JASP version 0.16.1) and are reported alongside.

Using simultaneous multiple regression, we examined how each variable affected Novel Truth Ratings for All items (i.e., both Delusion-Relevant and Neutral statements). The output from this analysis is reported in Table 3. It was observed that together, these predictors explained a significant proportion (39.5%) of the variance in Novel Truth Ratings, F (5, 90) = 13.40, $p < .001$, $R^2 = 0.40$. Familiarity was positively related to Novel Truth Ratings. Truth Status was not a significant predictor of truth ratings. Further, when analysed simultaneously with the other variables, Arousal, Accuracy, and Words were no longer significant predictors of Novel Ratings.

Table 4

Simultaneous regression output for key variables (All Novel Ratings)

	t	*b*	*SE*	*p*
1. Familiarity	7.08	.39	.06	*< .001
2. Arousal	1.18	.08	.07	.24
3. Accuracy	-.70	-.002	.003	.49
4. Words	1.00	.01	.01	.32
5. Truth Status	1.69	.20	.12	.10

Note. * = significant at the $p < .05$ level, uncorrected

Two more simultaneous regressions were then performed to investigate whether the two statement conditions behaved differently in terms of predicting Novel Ratings (depicted in Tables 5 and 6, below). These were conducted because we hypothesised that the discrepancy

in baseline truth ratings between Delusion-Relevant and Neutral statements resulted due to differences in average ratings across variables such as emotional arousal, familiarity, and statement length between the two conditions. It is possible that one or more of our predictor variables (e.g., Arousal) might explain variations between individual Delusion-Relevant item ratings, but not amongst Neutral ratings, for example.

The data was split into the two conditions and two multiple regressions were performed, again simultaneously entering the variables Familiarity, Arousal, Accuracy, Words, and Truth Status, to observe how they might affect Novel Truth Ratings. For the Neutral items only, we again found that our collection of predictors once again explained a significant proportion of variance in Novel Truth Ratings, $R^2 = 0.30$, $F\,(5, 42) = 5.08$, $p < .001$. Looking closer, however, only Familiarity was a significant predictor of ratings this time. Arousal, Accuracy, Words, and Truth Status were not significantly related to Novel Truth Ratings.

Table 5

Simultaneous regression output for key variables (Neutral Novel Ratings)

	t	*b*	*SE*	*p*
1. Familiarity	4.49	.42	.09	*< .001
2. Arousal	.53	.07	.07	.60
3. Accuracy	.71	.003	.004	.48
4. Words	-.64	-.01	.02	.53
5. Truth Status	-.56	-.09	.16	.58

Note. * = significant at the $p < .05$ level, uncorrected

When looking at Delusion-Relevant items only, our collection of predictors again explained a significant proportion of the variance in Novel Truth Ratings, $F(5, 42) = 11.71$, $p < .001$, $R^2 = 0.53$. Once again, Familiarity was a significant predictor of ratings. Further, and in contrast to the previous regressions, Truth Status was also a significant predictor of truth

ratings, as was Accuracy. Arousal and Words, however, were not significant predictors once the other variables were accounted for.

Table 6

Simultaneous regression output for key variables (Delusion-Relevant Novel Ratings)

Variable	*t*	*b*	*SE*	*p*
1. Familiarity	6.17	.40	.07	*< .001
2. Arousal	1.71	.16	.09	.10
3. Accuracy	-2.69	-.01	.004	*.01
4. Words	1.50	.02	.01	.14
5. Truth Status	3.55	.62	.17	*<.001

Note. * = significant at the $p < .05$ level, uncorrected

ANOVAs

Following the results reflecting a significant relationship between Familiarity and Novel Ratings, additional analyses were conducted to examine whether familiarity differed across each of our sub-categories (alternative medicine, pop-culture, etc). Mean familiarity ratings per item were entered into one way ANOVA with each statement sub-category constituting a level of the independent variable. Descriptive statistics are depicted in Table 7, in order of most to least familiar.

Table 7

Means and Standard Deviations for Familiarity in each Statement Category

Category	*N*	*M*	*SD*
Alternative Medicine	9	3.27	.96
Contamination	10	3.05	1.00
Persecution	10	2.81	.88
Science	10	2.70	.57

History	10	2.64	1.19
Conspiracy	10	2.42	.88
Trivia	10	2.26	.60
Geography	9	2.25	.45
Pop Culture	9	1.94	.46
Superstition	9	1.83	.44

Results of the one-way ANOVA revealed a large, significant main effect of Statement Category, $F(9,86) = 3.16$, $p = .002$, $\eta^2 = .25$. Post-hoc tests were performed using Tukey's honestly significant difference (HSD), but for consistency, q values were converted to t values (using JASP version 0.16.1). This revealed a significant difference in ratings of familiarity between Alternative Medicine and Pop Culture, $t(16) = 3.57$, $p = .02$, $d = 1.68$; Alternative Medicine and Superstition, $t(16) = 3.85$, $p = .008$, $d = 1.81$; and Contamination and Superstition, $t(17) = 3.33$, $p = .04$, $d = 1.53$.

Cronbach's alpha

To measure the internal consistency of each statement condition and its subcategories, a unidimensional reliability analysis was conducted. Cronbach's alpha was the measure of consistency (statistics presented in Table 8). Only data from Experiment 1 was used in this analysis, to ensure there was an equal number of participant responses for each item. Therefore, a total of 102 participants' initial (i.e., novel) responses to each item was included in the present analysis. A Cronbach's alpha of 0.7 or higher is considered to indicate good internal consistency (Tavakol & Dennick, 2011). Overall, the consistency across the whole scale (jointly considered) was good, and so too were the two primary categories (delusion-relevant and neutral) when considered separately. However, when the individual subcategories were considered in isolation, internal consistency fell away. This suggests some caution ought to be employed interpreting our findings at the subcategory level.

Table 8

Means and Standard Deviations for Novel Truth Ratings in each Statement Category, with Cronbach's alpha

Category	*N*	*M*	*SD*	α
All items	96	3.76	0.61	0.84
Delusion-Relevant	48	3.87	0.65	0.77
Alternative Medicine	9	3.70	1.10	0.53
Contamination	10	4.03	0.52	0.16
Persecution	10	4.19	0.46	0.64
Conspiracy	10	3.77	0.34	0.45
Superstition	9	3.60	0.54	0.71
Neutral	48	3.66	0.55	0.74
Science	10	3.74	0.51	0.21
Trivia	10	3.45	0.66	0.30
Geography	9	3.70	0.49	0.41
Pop Culture	9	3.60	0.39	0.52
History	10	3.82	0.67	0.38

Discussion

It was hypothesised that one or more of the key variables investigated in this chapter – familiarity, emotionality, and/or statement length – would share a positive relationship with novel truth ratings across Experiments 1 and 2 (Chapter 2). Results from the correlational analysis revealed several significant relationships between key variables, with familiarity emerging as the strongest predictor of novel truth ratings across analyses. Some significant correlations were also found between truth ratings and statement length, response accuracy, and truth status. However, no significant relationship was observed between novel truth ratings and emotional valence or emotional arousal. In addition, significant differences appeared in comparing ratings for emotionality and statement length across statement conditions, with

emotionality and statement length being significantly greater across delusion-relevant items than those that were neutral.

Novel ratings

It must be noted that Novel Ratings in fact did not return a significant difference between neutral and delusion-relevant items in this item-level analysis. While there was a measurable significant difference in ratings between statement conditions as reported earlier (a medium effect size difference between conditions was observed for both experiments in the previous chapter), this is not reflected in Table 2. However, this is due to the changes in the organisation of the data for the present analysis, where average ratings are compared across items rather than across each participant as in previous analyses (thereby markedly reducing our degrees of freedom and changing the variance profile). As the participant-by-participant analysis is a more accurate reflection of actual differences between conditions, the nonsignificant finding in Table 2 can be largely disregarded.

Familiarity

Familiarity and Novel Ratings shared the strongest bivariate relationship in the correlation matrix, indicating that Familiarity played a larger role in driving truth ratings than any other variable. Familiarity and novel truth ratings shared a positive correlation with a moderate effect size, indicating that higher levels of familiarity predicted higher truth ratings for novel items. This is in line with our initial predictions, suggesting that participants may have already encountered some of the items prior to the experiment. As a result, their first encounter with such items within the task may have elicited higher truth ratings than they would have offered if they were truly novel (i.e., a truth effect occurred before some items were repeated within the experiment).

As highlighted earlier in this chapter, there is evidence to suggest that subsequent repetitions can incrementally increase perceptions of truth (Hassan & Barber, 2021), meaning that we could still observe a truth effect occurring between first and second presentation of an item within our experiment – even if a participant's first encounter within the task already elicited an increase in truth from prior exposure to an item. While this seems to be the case, research shows this increase in truth judgement with subsequent repetitions to be logarithmic, with increases in truth being greatest between first and second encounter and becoming incrementally smaller with each repetition (Hassan & Barber, 2021). Essentially, this means that a truth effect can still be observed after more than one repetition, but that it will likely be a smaller effect. These observations fit well with the present data, whereby familiar items still elicited a truth effect after repetition, despite being the primary driver of higher truth judgements upon first presentation in the experiment.

As delusion-relevant items yielded higher initial ratings of truth and a smaller truth effect after repetition, it was posited that these items may have been more familiar than novel statements overall. The t-test was equivocal: no significant t-test was observed, but equally there was little evidence for the absence of an effect. Further, upon examination of Figure 7 (showing rankings of item familiarity from most to least familiar), it can be surmised that there was a clear tendency for delusion-relevant statements to be judged as higher in familiarity than neutral items. The top-heavy distribution seen in Figure 7 is likely important. To explain, suppose the relationship between familiarity and subjective truth is non-linear (as implied by the fact that repetition effects reduce in magnitude after the first repetition; Dechêne et al., 2010), or slightly non-transitive. If so, then a significant difference in familiarity between item types is not implied by their involvement in guiding initial truth ratings. This is because a small number of highly familiar items may exert an outsized effect on truth ratings, thereby leading

to significant differences between item types on truth ratings, while not being numerous enough to generate a significant difference between item types on familiarity ratings. It is for this reason that we hypothesized that this tendency for greater familiarity among delusion-relevant statements is the likely cause for greater truth ratings in Experiments 1 and 2 and should therefore be controlled for in subsequent experiments.

Emotionality

Large differences in ratings for both emotional valence and arousal were observed between delusion-relevant and neutral items in the current analysis. Results indicated a large difference in valence between statement types, whereby delusion-relevant items were judged to be more negative than neutral items. With emotional arousal, delusion-relevant items received far higher ratings than neutral items, suggesting a greater level of emotionality among statements with delusion-relevant content. These findings are consistent with our original hypothesis, whereby delusion-relevant items were predicted to be more negatively valenced and higher in emotional arousal than neutral items.

Despite these clear differences in emotionality between the two statement types, none of the regression analyses revealed an effect of emotionality on truth judgement, contrary to our initial predictions. This is interesting, given that emotionality differed greatly in both arousal and valence across delusion-relevant and neutral items, and this factor was intentionally manipulated in Moritz et al.'s (2012) original study. Yet differences in emotional content had no observable effect on the outcome of truth evaluation in the present study. Therefore, despite the strong rationale for predicting an influence of emotionality on judgements of truth, this hypothesis has not been supported by the present data.

While less important in the current findings, some additional significant relationships were observed between several key variables. Firstly, there was a small positive correlation

between emotional arousal and statement length (i.e., Words), with statements with higher levels of arousal tending to contain a greater number of words. Additionally, there was a small negative correlation observed between emotional valence and statement length, whereby statements with a more positive valence tended to contain fewer words than negatively valenced items. If we recall that delusion-relevant items contained, on average, twice the number of words per statement than neutral items, these findings could reflect a mere coincidence whereby delusion-relevant content elicited greater emotionality independently of the quantity of information provided. Alternatively, this finding could point to a relationship between quantity of information and its emotional impact, whereby increases in quantity of information affects a greater emotional response from its audience. Ultimately, however, we did not observe a clear relationship between emotionality and our primary dependent variable, subjective truth, and therefore focused on the unambiguous influence of familiarity (and accuracy, in the case of delusion relevant items).

Accuracy

In addition to familiarity, novel truth ratings also shared a small positive relationship with response accuracy. This indicates a small a tendency for true statements to be answered with higher accuracy than false statements overall. This finding might not point to an independent effect of accuracy or truth status on truth ratings but rather reflect the fact that some delusion-relevant items were explicitly known to be true by a subset of participants, thereby leading to higher familiarity ratings for those items and a positive relationship between objective truth and rated truth. Reassuringly, there was little evidence that this effect was widespread; response accuracy was not a significant predictor of novel truth ratings overall, or among the neutral items considered alone. This finding points again to the need for less

familiar, lesser-known items in the next experiment, particularly within the delusion-relevant statement condition.

Statement length

A small, positive relationship was observed between novel ratings and statement length, with statements containing more words tending to elicit greater average truth ratings. As delusion-relevant items contained, on average, twice the number of words per statement than neutral items, and delusion-relevant items were judged as being more true on initial encounter, this finding could reflect a role of information quantity and quality on truth judgements. Alternatively, this relationship might be spurious. Given that the correlation coefficient was small in our correlation matrix, and that Statement Length did not show up as a significant predictor of novel truth ratings in our simultaneous regression, it is more likely that statement length did not have a direct, independent influence on truth judgement.

Experiment 4

Based on the findings of the present analyses, it appears that familiarity was the most influential driver of initial truth ratings in Chapter 2. As a result, it was necessary to remove some of the most familiar items in the current list and replace them with lesser-known trivia claims. To ascertain which items would be replaced, each statement was ranked by familiarity and a second derived metric intended to be sensitive to pre-existing, explicit knowledge. This new metric – response confidence – refers to items that received the highest and lowest average response accuracy which, in addition to familiarity scores, points to previous knowledge of an item. Items with a very high response accuracy indicate that many people were able to correctly and confidently declare that item to be true (or false). Very low mean response accuracy indicates that many participants were sufficiently familiar with an item to rate confidently that

it was true (or false), but actually possessed inaccurate information about that statement. In either case, participants were likely familiar with that item, and their judgments of its truth would be disproportionately influenced by prior knowledge. Consequently, these items were replaced with items obtained from more obscure sources, so as to reduce the likelihood of them being highly familiar or known in advance.

Following development of the new list of 96 statements, an experimental survey was conducted to measure the performance of each statement. Specifically, it was important to identify whether the new list of delusion-relevant and neutral items returned similar average truth and familiarity ratings – i.e., the differences in ratings for truth and familiarity between neutral and delusion-relevant items should be nonsignificant. Once the two statement type lists are appropriately balanced, a final experiment investigating the influence of statement type on the truth effect can be conducted (see Experiment 5).

Method

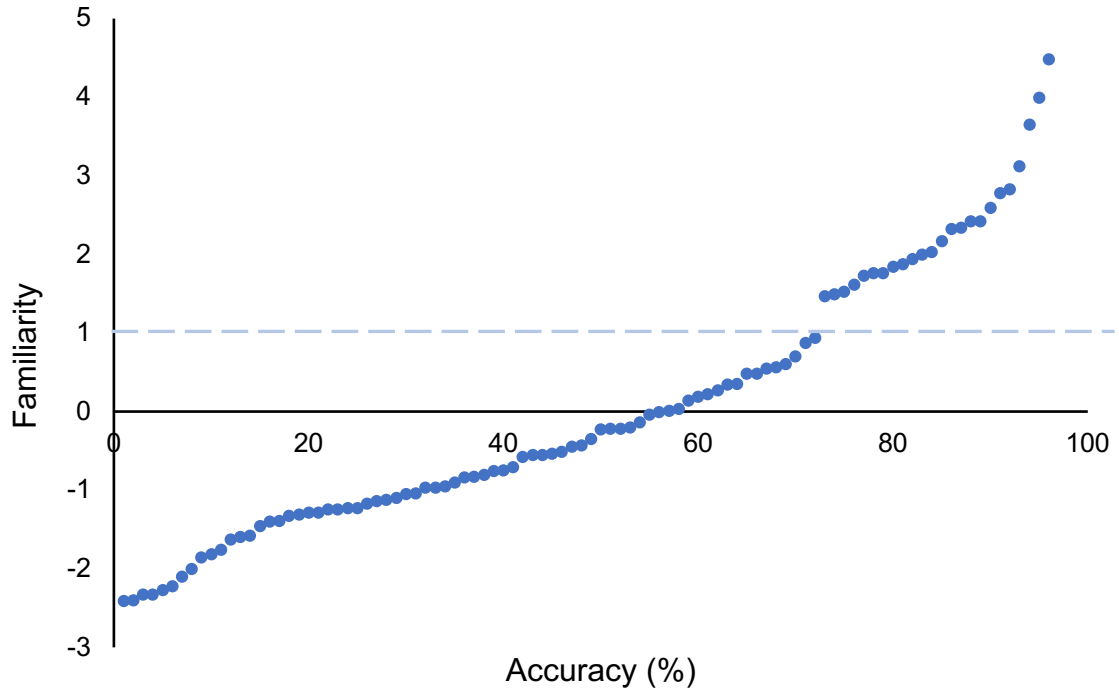
Materials

To rank items based on familiarity and response confidence, each statement received a weighted score for each variable. For response confidence, scores were determined by taking accuracy averages by items and recalculating as distance from the global mean. Therefore, those with a greater distance from the mean (i.e., with very high or very low response accuracy) ranked higher on the scale of response confidence. Once these weighted scores were determined, familiarity and confidence scores were transformed into weighted z-scores and added together. Each item was then ranked from highest to lowest based on its summed familiarity and response confidence score.

A total of 24 candidate items were identified as performing poorly for the purpose of the present research. The cut off point for items can be seen in Figure 8; all items with a composite score of 1 or higher are considered problematic due to high familiarity or explicit knowledge and will therefore be replaced.

Figure 8

All statements represented as a composite score based on familiarity (y axis) and accuracy (as a percentage score, x axis).



To confirm whether these items were indeed problematic, a t-test was performed to determine whether novel truth ratings still differed between statement conditions once those items were removed. When the 24 items with the highest familiarity-confidence composite score were removed from the analysis, there was no longer a significant difference in novel truth ratings between delusion-relevant and neutral items across Experiments 1 and 2, $t(70) = 1.46$, $p = 0.15$, $d = 0.35$.

These 24 items (15 delusion-relevant and 9 neutral; 21 true and 3 false) were then replaced with lesser-known trivia statements in the same proportion. New items were selected by accessing less popular or mainstream information sources. Delusion-relevant items were retrieved from websites closer to source material than we used when initially constructing the test items. Initially we sought interesting facts or urban myths from mainstream sources, whereas now we accessed technical materials, including food safety regulation documents, US defense research public-facing websites, and peer-reviewed journal articles. Neutral items were primarily sourced from the more difficult items in trivia games.

Participants

This study was approved by the Flinders University Human Research Ethics Committee (ID: 2671). Forty participants (24 female) were recruited online via Prolific Academic online recruitment platform (https://prolificacademic.co.uk/). Mean age of respondents was 34.33 years (SD = 13.22), and the majority (80%) were based in the United Kingdom, USA, Canada, New Zealand, and Australia. A further 15% were based in South Africa. All participants were listed as having English as a first language, and all received a payment of 2.25GBP for approximately 20 minutes of their time (6.75GBP/hour).

The experimental survey was created and distributed using Qualtrics survey engine (https://www.qualtrics.com), whereby participants were able to access the survey from their personal device. Qualtrics formats surveys to be easily undertaken on a computer, tablet, or mobile phone.

Procedure

Participants were informed that the present study aimed to examine how briefly presented words affect how people think, to better understand how and why people form rapid, intuitive judgements about the truth of a statement.

After consenting to participate in the study, participants were asked to provide their age and gender. They were then instructed to read the full new series of trivia statements (totalling 96 statements) and to separately rate each item on a 7-point Likert scale by:

1) how *true* (or false) they judged the content of the statement to be, and
2) how *familiar* (or unfamiliar) the statement was to them.

Participants completed the questionnaire in one sitting, with one statement presented per page, along with two Likert scales below each statement (one for Truth ratings and one for Familiarity ratings). Participants could respond to the two scales in any order, and their responses were untimed. The screen position of the scales was not randomized or counterbalanced (truth ratings were always in the upper location). The order of presentation of each statement was randomised between participants.

Results

Key variables

Average Familiarity and Truth ratings were calculated for each statement. The average familiarity and truth ratings for each category of statement (neutral and delusion-relevant) are depicted in Table 9.

Table 9

Descriptive statistics showing familiarity and truth ratings across Delusion-Relevant and Neutral items in Experiment 4

Variable	*N*	*M*	*SD*
Truth (All)	96	3.46	.51
Neutral	48	3.35	.48
Delusion-Relevant	48	3.56	.52
Familiarity (All)	96	2.65	.60
Neutral	48	2.67	.51
Delusion-Relevant	48	2.62	.68

Comparisons between statement types

An initial correlation showed a small, marginally significant positive relationship between truth ratings and familiarity ratings, $r = 0.20$, $p = 0.05$. Following this, two independent samples t-tests were conducted to compare truth ratings and familiarity ratings across statement types. For familiarity, the difference between delusion-relevant and neutral items was nonsignificant, $t(94) = 0.44$, $p = 0.66$, $d = 0.09$. This indicates that the new items performed better in terms of familiarity between statement types, supporting the removal of the previous items and inclusion of new ones. However, the second t-test revealed a significant difference in truth ratings between delusion-relevant and neutral items, indicating that, on average, delusion-relevant statements were judged as being more true than neutral statements overall, $t(94) = 2.04$, $p = 0.04$, $d = 0.42$ (see Table 8 for descriptive statistics). This indicates that, while the control for item familiarity across statements was successful, there remained a difference in baseline truth ratings for between statement types. Therefore, the current list needed further modification to ensure balance between neutral and delusion-relevant items.

Item list modifications

Rather than applying a second round of stimulus replacements, we instead modified the existing list of trivia statements using a more direct optimization method. Here, we simply removed several neutral items that had received high truth ratings, and delusion-relevant items receiving low truth ratings to directly force a balance on average truth ratings between neutral and delusion-relevant statements. To identify which items would be eliminated, the 6 delusion-relevant items with the highest truth ratings, and the 6 neutral items with the lowest truth ratings as measured in Experiment 4 were identified. This strategy was utilised in order to modify the average truth ratings of each category, so that the average neutral item rating was increased, and the average delusion-relevant rating was decreased. The decision to remove 12 items meant that the number of items per list could still be counterbalanced appropriately, including an equal number of items per statement type. Analyses were run with this new item list, totalling 84 statements (42 true, 42 false; 42 neutral, 42 delusion-relevant), to ascertain whether truth ratings were balanced between statement types before running the final truth effect experiment. This new set of items is listed in Appendix B.

Table 10

Descriptive statistics for truth and familiarity ratings across statement types after twelve items removed

Variable	*N*	*M*	*SD*
Truth (All)	84	3.47	.41
Neutral	42	3.48	.36
Delusion-Relevant	42	3.46	.47
Familiarity (All)	84	2.59	.53
Neutral	42	2.67	.43
Delusion-Relevant	42	2.50	.60

An independent samples t-test confirmed there were no significant differences in familiarity ratings between neutral and delusion-relevant statements for the new list, $t(82) = 1.56$, $p = 0.12$, $d = 0.34$. Further, another independent samples t-test revealed the difference in novel truth ratings between neutral and delusion-relevant items was no longer significant, $t(82) = 0.29$, $p = 0.78$, $d = 0.06$. This indicated that the current list of 84 items was now balanced between statement types and was therefore suitable to utilise for the final truth effect experiment. Further, a bivariate correlation between truth and familiarity confirmed these variables do not share a significant relationship, $r = 0.12$, $p = 0.28$.

Discussion

A two-step refinement process arrived at a slightly shorter list (84 items versus 96) that was balanced on both familiarity and novel truth ratings across neutral and delusion-relevant categories. However, the assertion of balance was fundamentally made on a non-significant, post hoc analysis of the second study, after manually removing the items needed to achieve balance. For this reason, it was important to test this item set in a new sample of participants to determine whether the two lists were actually balanced, and secondly, whether any difference in truth effect magnitude would be evident between neutral and delusion-relevant categories, holding constant the initial truth ratings for both categories.

Experiment 5

This experiment shared the same aims, hypotheses, method, and design as Experiment 2. Specifically, we sought to identify whether delusion relevant content modified the magnitude (or direction) of the truth effect. The only differences were that this experiment used the modified, 84-item stimulus list developed in Experiment 4, and it was tested on a new participant sample.

A note on the present hypothesis: It is possible that we may see a higher truth effect in delusion-relevant items even in a non-clinical sample, if we consider the prevalence of belief in conspiracy theories among the general population. The content of such strongly held beliefs, while not delusions in the clinical sense, share similar themes to those in delusions (in fact, one of the subcategories of delusion-relevant statements is 'conspiracy beliefs'). In addition, we know that belief in fake news has proliferated among the population – surviving regardless of implausibility (Fazio et al., 2019). We therefore may find that people, in general, respond to the effect of fluency more greatly in the face of delusion-like content.

Method

Participant recruitment

This study was approved by the Flinders University Human Research Ethics Committee (ID: 2671). One hundred participants (45 female) were recruited online via Prolific Academic online recruitment platform (https://prolificacademic.co.uk/). Only subjects who reported English as their first language were recruited, and participants received a payment of 4.20GBP for approximately 30 minutes of their time (8.40GBP/hour). Mean age of respondents was 37.23 years (SD = 12.04), and the majority (76%) were based in the United Kingdom, USA, Canada, New Zealand, and Australia. A further 16% of participants were based in South Africa.

The experimental survey was created and distributed using Qualtrics survey engine (https://www.qualtrics.com), whereby participants were able to access the survey from their personal device.

Procedure

The present experimental procedure was identical to that of Experiment 2 in Chapter 2, but with 12 fewer trivia statements. After consenting and providing initial demographic data,

participants completed the O-LIFE and the CAPE-42. Participants were then sequentially shown 56 statements and were asked to indicate how interesting (or uninteresting) they found each statement, using an identical onscreen Likert scale to that used in Experiment 2. This was followed by three questions relating to participants' perception of time spent on the experiment so far, to break up the first from the second experimental phase. In the second phase of the experiment, participants were asked to rate how true (or false) they believed each statement to be. This phase consisted of another 56 statements, half of which were repeated from phase one, and half of which were novel. Again, half (n = 28) of these statements were delusion-relevant (vs neutral), and half were true (vs false).

Overall, each participant evaluated the truth and/or interestingness of a total of 112 statements, with 28 occurring twice over the course of the experiment. The order of presentation of items was randomised, and the allocation of each item to either repetition or once-only, and to occur in the first and/or second phase of the experiment, was counterbalanced across participants.

Results

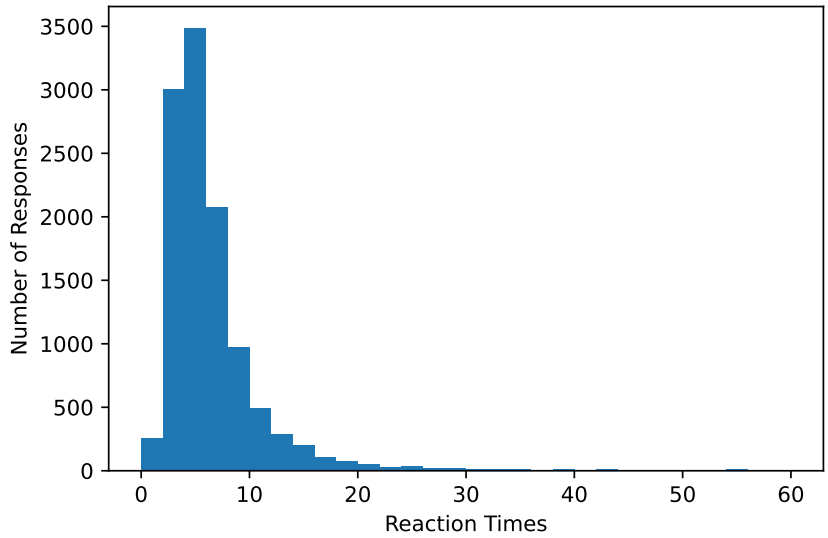
Response times

Participant responses taking less than 1 second or more than 22.5 seconds were removed from the final dataset. Responses that took less than 1 second were removed for the assumption that corresponding items were not able to be read before participants selected their response. Responses above 22.5 seconds were more than one standard deviation above the mean response time and were removed to control for participants potentially finding the answer to the trivia claim before responding. See Figure 9 for histogram showing response times for all truth and interestingness evaluations.

While we removed all responses above 10 seconds in Experiment 2, unfortunately, this was not possible in the current Experiment without removing a much larger proportion of the overall data; this group of participants had a larger, longer positive tail in their response latency distribution than did our first sample (see Figure 9). It remains unclear why this was so. However, because response latency is solely used to ascertain compliance with the task, and is not the subject of inferential analyses, we did not explore this issue further. Additionally, all inferential conclusions from earlier experiments held with or without the application of these criteria. For consistency, a criterion was applied that was numerically more lenient (>10s) but was itself a common method of cleaning: all responses longer than one standard deviation above the mean were removed. Overall, 2.67% of all truth rating responses were removed from the final dataset, leaving a total of 10,901 item responses for analysis. Again, any participant with more than 20% of data removed were to be excluded, however no participant met this criterion. Once again, removal of this data did not alter the pattern of significance observed.

Figure 9

Means (and standard deviations) for ratings for combined, neutral, and delusion-relevant items across training (i.e., 'interestingness' evaluation) and test (i.e., 'truth' evaluation) phases.



Delusion-proneness measures

The Short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE). Scores over most categories were similar to previously documented population norms (Mason, Linney, & Claridge, 2005; Fonseca-Pedrero et al., 2015). For each subscale, we inferentially compared the obtained scores of our sample against the standardization sample using z-scores. The average score for unusual experiences was 3.56 (SD = 2.72), with a z-score of 0.07, p = 0.47. For cognitive disorganisation, the mean was 5.34 (SD = 3.00), z-score = 0.32, p = 0.37. Average scores for introvertive anhedonia and impulsive nonconformity were 3.76 (SD = 2.44), z-score = 0.65, p = 0.26, and 2.94 (SD = 1.95), z-score = 0.17, p = 0.43, respectively.

The Community Assessment of Psychic Experiences (CAPE-42). For the CAPE, the corresponding frequency response for missing distress response data fields were removed. The average weighted score for frequency of positive symptoms was 1.41 (*SD* = 0.36), while the weighted average for negative symptom frequency was 2.10 (*SD* = 0.50), and the average for depressive symptom frequency was 2.06 (*SD* = 0.54). Z-scores were again computed to compare our population means with the healthy population CAPE-42 score averages reported by Moritz et al. (2012). This revealed that the weighted frequency scores from each population are comparable across positive (*z*-score = 1.11, *p* = 0.13), negative (*z*-score = 0.92, *p* = 0.18) and depressive (*z*-score = 0.61, *p* = 0.27) symptom frequency.

The truth effect

A paired samples t-test was conducted to examine whether the two statement types elicited similar average truth ratings at first encounter (i.e., when novel). This confirmed that the two statement lists were not significantly different and were indeed balanced, $t(99) = 1.08$, $p = 0.28$, $d = 0.11$. Calculation of the between-items truth effect was then obtained by subtracting truth ratings for repeated items from ratings for novel items in the second half of

the survey. As in Experiment 2, no within-items truth effect was obtainable due to participants rating items in the first half of the experiment by interestingness, rather than by truth.

Figure 10

Means (and standard deviations) for ratings for combined, neutral, and delusion-relevant items across (a) training (i.e., 'interestingness' evaluation) and (b) test (i.e., 'truth' evaluation) phases.

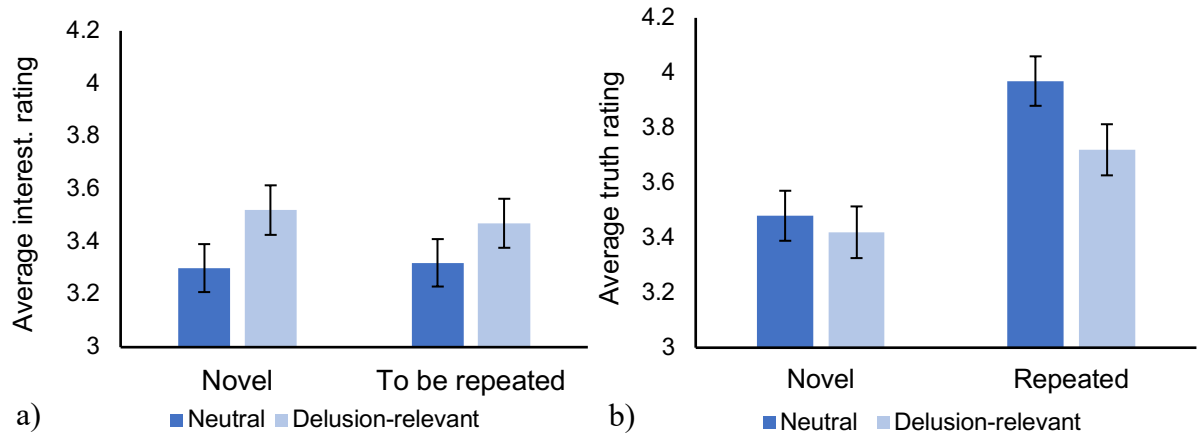


Figure 10 shows average truth ratings for novel and repeated statements across statement types. A clear difference in average truth ratings exists between novel and repeated items, signifying a significant effect of repetition on truth. There also appears to be a significant difference in ratings for repeated items between neutral and delusion-relevant statements (seen on the far right), in contrast to ratings for novel items.

A 2 (repetition: repeated, not repeated) x 2 (statement type: neutral, delusion-relevant) repeated measures ANOVA was carried out to investigate whether a difference in truth ratings indeed occurred between novel and repeated statements in the second half of the experiment (i.e., if there was in fact a significant between-items truth effect). Consistent with our initial predictions, and the findings of the previous chapter (Experiment 2), a significant main effect of repetition on truth ratings was observed, indicating a large effect, $F(1,99) = 39.47$, $p < .001$, $d = 0.92$. In addition, the main effect of statement type was significant, showing a small-

medium size difference in mean truth ratings between delusion-relevant and neutral statements, $F(1,99) = 12.87$, $p < .001$, $d = 0.34$. The interaction between statement type and repetition was also significant, revealing a small effect, $F(1,99) = 7.63$, $p = .007$, $d = 0.20$.

Post-hoc t-tests were conducted to take a closer look at the initial ANOVA findings. The prediction was that the size of the truth effect would be greater for delusion-relevant items than for neutral items. However, the truth effect was in fact larger for neutral items than delusion-relevant items. The truth effect (i.e., difference in truth ratings between novel and repeated statements) for each category was large and significant when considered in isolation: neutral items, $t(99) = 6.43$, $p < .001$, $d = 0.64$, and delusion-relevant items, $t(99) = 4.46$, $p < .001$, $d = 0.45$. This finding is in exact opposition to our initial hypothesis, which predicted a larger truth effect among delusion-relevant than neutral items.

Moritz et al. (2012) items and truth

Unfortunately, it was not possible to isolate and test the truth effect and schizotypy for items taken from Moritz et al. (2012) for this experiment, given that a significant proportion of those original items did not survive the modification process carried out in Experiments 3 and 4. Only 4 of Moritz et al.'s delusion-relevant items remained in the present experiment, and given that only one third of participants experienced each of these items as repeated, there is insufficient data to make appropriate inferences regarding their efficacy here.

The effect of delusion-relevant item category on truth

Within the delusion-relevant item type, there were 5 different categories: Persecution, Contamination, Conspiracy, Supernatural Belief, and Alternative Medicine. Due to the final number of items required for this experiment, it must be noted that there was an unequal number of items in each category. The number of items per category was either 9 or 10.

Analyses explored the pattern of truth evaluation within each category, to see whether there were significant differences between them. As some categories (e.g., persecution) were arguably more related to clinical delusions (i.e., delusions of persecution), than others (e.g., alternative medicine), we hypothesised that items relating to persecution would elicit larger truth ratings and a larger truth effect than other categories in the delusion-relevant condition.

Table 11

Magnitude of the truth effect for each delusion-relevant item category and all neutral items in Experiment 5

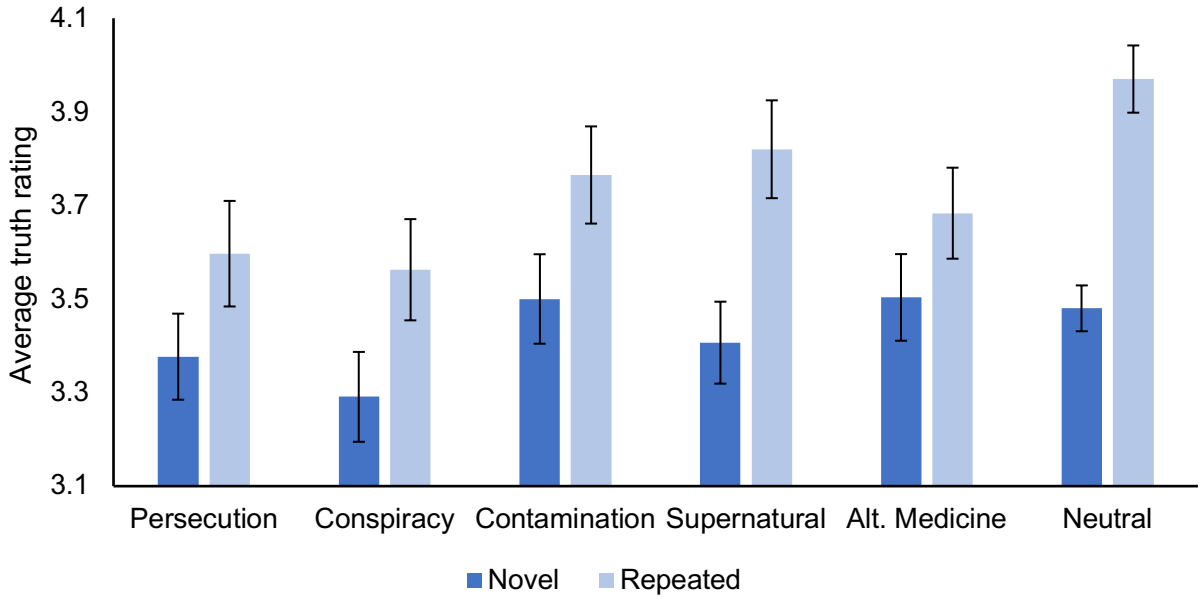
Item category	*t*	*p*	*Cohen's d*
Alternative medicine	2.04	*.044	.20
Supernatural	3.92	*<.001	.39
Conspiracy	2.12	*.035	.21
Persecution	1.66	.101	.17
Contamination	2.12	*.037	.21
Neutral (all items)	6.43	*<.001	.64

** Denotes p values indicating significance*

A repeated measures ANOVA was conducted with two factors: item category (Persecution, Contamination, Conspiracy, Alternative Medicine, Supernatural Beliefs, and Neutral), and repetition status (repeated, novel). This revealed a significant main effect of item category, $F(5,485) = 2.95$, $p = .012$, $d = 0.25$ (see Figure 11). Again, a difference between repeated and novel statements was significant and clearly observable, $F(1,97) = 23.21$, $p < 0.001$, $d = 0.39$. The interaction between item category and repetition was not significant, however, $F(5,485) = 1.40$, $p = 0.22$, $d = 0.13$. This finding differs from that of the previous chapter, where there was a significant interaction observed between item category and repetition status.

Figure 11

Average truth ratings (with standard error bars) comparing novel and repeated items across Delusion-Relevant item categories and all Neutral items.



Discussion

A large, significant effect of repetition on truth was observed across all items. This means that our list of statements effectively induced a repetition truth effect. There appeared to be a clear difference in truth ratings between statement types, indicating that neutral and delusion-relevant content affected truth evaluation differently. Finally, and most importantly, the magnitude of the truth effect differed significantly between statement types, indicating that repetition affected truth evaluation differently for neutral and delusion-relevant content. The direction of this effect was opposite to our initial prediction.

The truth effect

In the present chapter, we sought to understand how processing fluency (as manipulated by repetition) might differentially affect the way we evaluate the truth of

statements with different types of content, namely, delusion-relevant versus neutral trivia claims. After controlling for numerous extraneous variables, modifying content, and removing poorly performing items, findings from Experiment 5 support the hypothesis that people evaluate the truth of different types of information differently. Indeed, results indicated that average truth ratings, as well as the magnitude of the repetition truth effect, differed between items that are related to delusional content and those that are neutral.

However, the direction of the findings is in direct contradiction to our initial hypothesis: *neutral* content elicits a greater effect on truth with repetition than content related to delusions. These findings echo and strengthen the trend of results seen in Experiment 2, where neutral items also elicited a greater truth effect than delusion-relevant items, but where unmatched baseline truth ratings rendered any conclusions about content type and truth evaluation ambiguous. As can be seen in Figure 10, Experiment 5 paints a much clearer picture of different types of content and their effect on perceptions of truth.

While unexpected, there are some potential explanations for these findings. For example, perhaps people in the general population process and interpret delusion-relevant content with more caution and are therefore more careful when it comes to judging their veracity after repetition. As discussed at the beginning of this chapter, stimuli with a negative valence have been demonstrated to promote greater and deeper processing and more deliberate evaluation than positive stimuli (Bohner et al., 1988; Lewicka, 1997); findings which seemingly contradict the claims that negatively valenced information is more fluently processed. It is possible that items relating to delusions (e.g., themes of persecution or conspiracy) may have elicited a greater depth of processing than neutral items due to their negative emotional valence, thereby reducing the involvement of intuitive, perceptual judgment processing thought to underlie the truth effect. The emotionality of this content could

have had an additional influence on processing and/or reasoning in a way we did not anticipate. Rather than increasing the perception of validity for delusion-relevant items with repetition, for example, this more elaborate processing and careful consideration may have diminished the size of the truth effect. This may be due to deeper encoding at first encounter of such items, so that people could better remember their initial perception of the truth of those items and remain more consistent (similar to what we observed in Experiment 1). Indeed, delusion-relevant items elicited significantly longer response times than neutral stimuli in Experiment 5, $t(83) = 7.64$, $p < .001$, $d = 0.83$. This would support the above account. However, it must also be explicitly stated that the differences in response times between the two statement types is likely to be partially attributable to the differences in average statement length (remembering that delusion-relevant items contained, on average, more words than neutral items).

Alternatively, people may have simply been more suspicious of the validity of delusion-relevant items and remained more conservative in their evaluation throughout the experiment – after all, participants were informed that half of the statements were false and may have perceived that those claims that were more negative, emotional, or controversial were more likely to be false. Given that people are not typically given a disclaimer stating that information they are receiving is only 50% true in the real world, it follows logically that people may have been more suspicious of repeated emotional content in the current experimental setting than they might be in less constrained, more open-ended experimental or applied settings. People may also have felt more conservative about assigning truth to delusion-relevant items due to the stigma attached to believing in such claims (i.e., those which contradict the societal beliefs), although the higher ratings at first presentation tend to contradict this explanation. This interpretation is necessarily offered speculatively because we predicted an effect in the opposite direction.

Delusion-relevance and item category

As our delusion-relevant statements contained different subcategories of thematic content, analyses were conducted to test whether these subcategories performed differently from one another and from neutral statements. As previously noted, some item categories contained claims that were more related to themes of clinical delusions, such as persecution (e.g., "The resolution of Earth observation satellites is so high that it is possible to read car license plates from a height of 300km"), while others were arguably less related, such as alternative medicine (e.g., "Reiki helps with symptoms of asthma, chronic fatigue, menopausal symptoms and arthritis"). Therefore, it was predicted that items in the subcategory Persecution would elicit a greater truth effect than neutral statements and other delusion-relevant item categories.

The results indicated there was a significant difference between item categories, but no significant interaction effect between item category and repetition. This indicates that, while different types of delusion-relevant statements elicited significantly different truth ratings overall, they did not appear to differentially affect the magnitude of the repetition truth effect. Surprisingly, the Supernatural item category elicited the largest truth effect among delusion-relevant items, while Persecution was the only item category to not elicit a truth effect at all (see Table 11). This suggests that people in the present experiment may have been more suspicious or uncertain when it came to claims that were thematically related to clinical delusions. It will be interesting to explore this difference in truth evaluation between delusion-relevant subcategories further in the next chapter, when we investigate the impact of schizotypy (and other individual difference variables) on the truth effect. For example, we may see a different pattern of results emerge once we consider those with high versus low schizotypy and the relative size of the truth effect in each group.

General discussion

The findings of this chapter can be summarised by addressing each of the three experiments conducted separately. The chapter commenced with an investigation (Experiment 3) of which variable/s predicted the baseline inequality in novel truth ratings observed between delusion-relevant and neutral items in Experiment 2. While the findings suggested that more than one key variable shared a relationship with judgements of truth, familiarity was identified as the predominant factor responsible for the discrepancy in novel ratings. Whether this familiarity originated in explicit knowledge of previous encounters with some claims, a sense of recognition of the thematic content in some items, or simply familiarity with the component parts of the claims, we cannot determine from the present findings. Interestingly, Experiment 3 found that neither emotional arousal nor valence appeared to significantly affect truth judgements in Experiments 1 and 2, despite a clear large difference in emotionality between neutral and delusion-relevant statements. This suggests that the emotionality of content does not influence initial judgements in a standard truth effect experiment, at least within the items and methods used in the present protocol. However, while we have not observed an effect of emotionality on novel truth in this case, we must keep in mind that the participants who judged the emotionality of items were from a different sample to those participants who completed the truth effect experiment. As such, we cannot definitively rule out the potential role of emotionality (or other variables, for that matter), on truth ratings in Experiments 1 and 2. Further, while the two statement conditions were not perfectly matched in the average length of statements, this did not appear to influence truth, as the variable did not emerge as a significant predictor of truth ratings in the regression analysis.

Experiment 4 sought to test a new list of delusion-relevant and neutral statements after 24 items with high familiarity were replaced with lesser-known claims. An analysis of truth and familiarity ratings provided by a small sample of participants revealed that, while now matched in familiarity, the two lists still differed significantly in truth ratings. From this, the list was further modified by removal of the highest-rated delusion-relevant and lowest-rated neutral items, revealing a finalised list of eighty-four trivia statements that were successfully matched in familiarity and truth.

This 84-item balanced list was used in Experiment 5. This experiment observed a surprising result: neutral content elicited larger truth effects than delusion-relevant content. These findings contradict those reported by Moritz et al. (2012). Given that the present work represents an adaption (and enhancement) of the pilot study conducted by Moritz et al., how then can we understand the directional reversal of the findings? The limitations acknowledged by the authors themselves present some clues, but a detailed discussion of this point will be held to our concluding chapter (Chapter 6).

Instead, we now turn to our second motivating question, which concerns whether differences amongst people leave them more or less vulnerable to the truth effect. Specifically, in the next chapter we consider whether schizotypy predicts the magnitude of the truth effect, and separately, whether it predicts belief in delusion-relevant content generally. While our primary interest lies in the sub-clinical analogues of the positive symptoms of schizophrenia, we additionally examined the other subscales of schizotypy, and adjacent constructs (such as delusion-proneness). In that context, it remains possible that any effect of belief content may interact with individual-level differences to cumulatively modify the truth effect. That is, content effects may only be present in those

with high levels of the positive aspects of schizotypy, and the present analyses would be blind to such subtle interactions. These hypotheses will be explored in the next chapter.

Chapter 4: Individual Differences in the Truth Effect

Not everyone develops high levels of conviction for any given idea. This is evident in the general population, where there is variation in the level of conviction held for any one belief, both between and within individuals. Most of the time, ideas held with high conviction have a reasonable basis – for example, if there is adequate objective or social support for adopting that belief (as in the case of scientific consensus, or religious or cultural beliefs; Coltheart et al., 2011; Jones & Watson, 1997). Sometimes, however, high levels of conviction are held for an idea that is not supported by objective evidence, nor shared by one's community, and is considered by most to be unfounded. In the extreme, conviction in these types of beliefs can lead to significant social or occupational dysfunction in a person's life and then may be classified as a delusion (American Psychiatric Association, 2013).

Conviction is a crucial feature used to determine a diagnosis of clinical delusions in psychotic disorders such as schizophrenia (American Psychiatric Association, 2013). Levels of conviction are commonly used to distinguish between overvalued ideas and delusional beliefs, as delusions are held with greater conviction and thereby exhibit greater inflexibility (Coltheart et al., 2011). In addition, delusions often (though not always) have an abrupt onset, with the belief and its conviction appearing rather spontaneously (Jones & Watson, 1997). These distinctions are important in differentiating between psychotic (e.g., schizophrenia) and nonpsychotic (e.g., mood) disorders, as overvalued ideas as seen in nonpsychotic disorders are less fixed and tend to develop over time (Mullen & Linscott, 2010). Cognitive treatments for delusions often involve working to reduce high conviction by employing techniques developed to promote analytical reasoning and challenge belief plausibility (e.g., Alford & Beck, 1994). While somewhat effective (Haddock et al., 1998), these strategies are often insufficient to

eliminate delusions, which are by definition resistant to change and often immune to disconfirming evidence (Moritz & Woodward, 2006).

A core hypothesis of this thesis is that delusions are related to an overreliance or disruption in intuitive reasoning processes (Balzan et al., 2012; McLean et al., 2016; Ward & Garety, 2019). To this point, we have explored and documented one intuitive mechanism (i.e., the truth effect) in which belief conviction is affected by mere repetition. This might be one mechanism by which sub-cultures maintain their unique, subjective beliefs, which is then also supported by a general social pressure to conform (Ellinas et al., 2017; Holland et al., 2003). But could this mechanism also promote the formation of idiosyncratic, empirically unsupported beliefs in a sub-culture consisting of a single person? If so, are some people – for example, those who are more prone to developing delusions – simply more susceptible to relying on intuitive reasoning than others? The present chapter will seek to answer this question by exploring the relationship between the magnitude of the truth effect and individual differences in trait schizotypy and delusion-proneness.

The feeling of conviction

The feeling of conviction is an involuntary metacognitive sensation (Burton, 2009) that we use as a cue for evaluating the truth or validity of a given piece of information (Unkelbach & Greifeneder, 2013). This metacognitive sensation can influence judgements whether one is consciously aware of the sensation of something 'feeling true' or not; and regardless of whether one is able to consciously ascertain the source of the feeling (e.g., because it is in an easy-to-read font, because one has heard it before, or because one has an ideological incentive to believe in said information; Dechêne et al., 2010). As such, the use of feelings of knowing or conviction in the formation of judgements and beliefs is a function of intuitive reasoning,

because it does not require nor arise via explicit, deliberative thinking processes (Kahneman, 2011).

Feelings of conviction can motivate attachment to beliefs that we know have no empirical evidence to uphold them, such as in the case of supernatural phenomena, superstitions, and clinically held delusions. For example, listening to one's own heartbeat might do nothing to dissuade a person who believes they are clinically dead, as in the case of patients with Cotard's delusion (Berrios & Luque, 1995; Burton, 2009). Similarly, a person with Capgras delusion may remain convinced an antique piece of furniture has been replaced with a replica without being able to identify any objective differences in its appearance ("'I admit that it is physically impossible that the desk has been replaced. But it has [been replaced]… *I know what I know*"'; Burton, 2009, pp. 16-17). While the phenomenon of erroneous conviction is well-known and documented, we do not yet understand which cognitive factors influence its development in some individuals, while not in others.

We know that one source of feelings of conviction is the degree to which ease-of-processing occurs, as is exemplified by the truth effect phenomenon (Dechêne et al., 2010). Processing fluency, as we have previously described, is a metacognitive sensation that is experienced as the relative ease or difficulty in processing a piece of information (Unkelbach & Greifeneder, 2013). It can be manipulated by repetition (Jacoby & Dallas, 1981), colour contrast (Reber & Schwarz, 1999), and font type (Unkelbach & Greifeneder, 2013), as well as by manipulations such as rhyming (McGlone & Tofighbakhsh, 2000), and semantic priming (Kelley & Lindsay, 1993). The experience of processing fluency influences a range of judgements, including liking (Westerman, Lanska, & Olds, 2015), familiarity (Whittlesea, 1993), confidence (Alter et al., 2007), and source credibility (Begg et al., 1992; Henkel & Mattson, 2011), depending on the type of judgement required. In judgements of veracity,

processing fluency is experienced as a subjective feeling of knowing or conviction, which is then utilised intuitively as a cue for truth.

The core question of this thesis is whether reliance on intuitive processes – such as the experience of processing fluency – differs between individuals and types of content to the degree to which it could help predict or explain the development of erroneous conviction in clinical delusions. To this point, our investigation has explored the role of different types of thematic content on the development of belief in an idea (e.g., trivia claim) following repetition, through the lens of the truth effect. We have observed that delusion-relevant content does not preferentially elicit intuitive reasoning when assessing veracity; if anything, non-delusional content appears to evoke a greater effect of repetition on truth. This chapter will address the other half of our question: that is, are there individual differences within the population that help explain the likelihood or degree to which one's belief in an idea will be augmented over time? Specifically, we wish to determine whether subclinical personality traits associated with schizophrenia – namely, schizotypy and delusion-proneness – predict the degree to which repetition increases perceptions of truth in individuals in the general population.

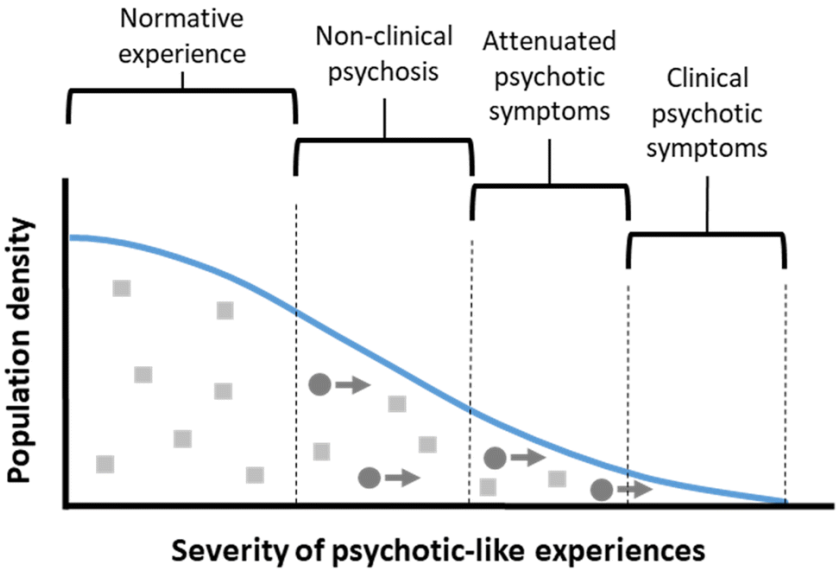
First, we will explore the fully dimensional model of psychosis, which posits that psychosis, psychosis risk and nonclinical psychosis-like traits and experiences exist along a continuum (Cowan & Mittal, 2021; Grant et al., 2018). This theory is well-supported across the literature, and enables researchers to explore symptoms and features of psychotic disorders using a general population sample. We will then examine the research to date surrounding individual differences in the repetition truth effect. While the truth effect has been largely demonstrated to be robust against differences between individuals, there is preliminary evidence that schizotypy or psychosis-proneness influences truth judgements for repeated claims.

The psychosis continuum

Of course, in exploring the relationship between traits like schizotypy and the truth effect, we are not directly testing the effect of repetition on conviction in clinical delusions. However, there is convincing evidence to support the notion that psychotic symptoms exist on a continuum within the general population (see Figure 12). While the prevalence of psychotic disorders in the general population is low – affecting approximately 1.5% of individuals – the prevalence of psychotic symptoms is more common (Hinterbuchinger & Mossaheb, 2021; van Os et al., 2000). In fact, a recent survey of 34,653 US citizens reported that more than one quarter (26.69%) of participants had encountered some form of psychotic-like experience (PLE; Bourgin et al., 2020). Earlier reports have suggested a similar prevalence among the US population (e.g., 28% in the National Comorbidity Survey; Kendler et al., 1996), while some European reports have been slightly lower at 17.5% (e.g., van Os et al., 2000; Spauwen et al., 2003). Furthermore, there has been some evidence to suggest that the presence of these psychotic symptoms persist over time (Rössler et al., 2007), suggesting that such experiences are more reflective of traits, rather than short-term states or temporary conditions. As such, it is widely held that the existence of psychosis or psychosis-like personality traits occur along a spectrum, with those with imperceptible levels of psychotic experiences on one end, and those with psychotic disorders such as schizophrenia on the other. This is referred to as the fully dimensional model of psychosis and is related to the nonclinical personality trait known as schizotypy (Nelson et al., 2013).

Figure 12

Cowan & Mittal (2021) – Psychotic-like experiences in the general population



Note: A hypothetical representation of the distribution of psychotic-like experiences in the general population in a fully dimensional model of psychosis and assuming a continuous distribution. From "Three types of psychotic-like experiences in youth at clinical high risk for psychosis" by H. R. Cowan and V. A. Mittal, 2021, *European Archives of Psychiatry and Clinical Neuroscience, 271*(4), p. 734. Copyright 2020 by Springer-Verlag GmbH Germany, part of Springer Nature. *Reproduced with permission from Springer Nature.*

Schizotypy

Schizotypy refers to a constellation of personality traits relating to key aspects of psychosis, including delusion-proneness (Meehl, 1962). It assumes that psychosis can be broken down into different categorical features, and that the presence of each of these features exists on a continuum. More specifically, according to Claridge (1985), schizotypy consists of a range of personality traits which reflect variation in cognitive style and perceptual experiences, that are sustained across an individual's lifetime, and which are normally distributed within the general population. The traits comprising schizotypal personality are neither diagnostic nor predictive of psychopathology, though presence of these traits (particularly in higher levels), in combination with various biological and psychosocial factors,

is thought to influence transition to psychiatric illnesses such as schizophrenia (Grant et al., 2018). Therefore, while schizotypy is conceptualised as a necessary condition for the development of psychotic illness, it does not sufficiently account for the onset of such illness in any one individual.

The origin of schizotypal personality traits are considered to arise from a combination of polygenetic and environmental factors, resulting in several observable endophenotypes which differ in degree and presentation across individuals (Grant et al., 2018). Therefore, the presence of these traits can range from nonclinical to clinical thresholds and can be useful in indicating those who are at clinical high risk (Cowan & Mittal, 2021). Due to the conceptualisation of schizophrenia as a spectrum disorder, measurement of schizotypal personality traits is often used in research investigating schizophrenia and schizophreniform disorders, posing as a useful proxy for psychosis-proneness, risk, and symptomatology (Mason, 2015).

It must be noted that some research (though not all) treats schizotypy as fundamentally independent of psychosis-proneness, due to the discrete differences between schizotypal personality disorder and schizophreniform disorders as described in the current DSM (American Psychiatric Association, 2013; Grant et al., 2018). Such observations highlight the need for conceptual clarity and consensus within research surrounding psychotic-like experiences, schizotypy, and psychosis-proneness (Grant et al., 2018; Hinterbuchinger & Mossaheb, 2021). Some measures of schizotypy may be more reflective of psychosis risk specifically, and in higher scores might even be indicative of a psychotic disorder (e.g., CAPE-42; van Os, Verdoux, & Hanssen, 1999), while others may be more reflective of schizotypal personality as its own distinct phenomenon (e.g., O-LIFE; Mason & Claridge, 2006). For this reason, the present investigation will use two different measures of schizotypy – the CAPE and

the O-LIFE – which were composed for different purposes and using different (i.e., clinical vs nonclinical) samples in their development.

Although we used a measure from either end of the schizotypy-as-trait versus schizotypy-as-psychosis-proneness continuum, the terms schizotypy and psychosis-proneness are often used interchangeably in research of psychotic disorders and psychosis risk in the general population. Further, there is no discrete difference between schizotypy and psychosis-proneness measures in terms of their capacity to effectively measure both clinical risk and nonclinical trait variation (Mason, 2015). Fundamentally, categorical dimensions of schizotypy and psychosis-proneness remain essentially the same; particularly in their acknowledgement of anomalous experiences, which play a key role in the development of delusional beliefs. This is crucial, given that the measure of delusion-proneness is the key focus of the present research. Furthermore, these terms are more often treated synonymously, and large correlations exist between several scales favouring either term in their description (Mason, 2015). For this reason, the terms ‘schizotypy’ and ‘psychosis-proneness’ will be used interchangeably in the present investigation.

The most common approach to conceptualising schizotypy or psychosis-proneness is based on the idea that psychotic symptoms can be broken down into three key subcategories: positive, negative, and disorganised symptoms (American Psychiatric Association, 2013). Measures typically capture positive symptoms as reflected by aberrant sensory, perceptual, and cognitive experiences (including magical ideation, delusional thinking, and hallucinations); negative symptoms as reflected by phenomena such as flat affect, anhedonia, and avolition; and cognitive or disorganised symptoms such as odd behaviour, and confused or incoherent speech (Kwapil et al., 2018). These three symptom categories can be observed across most

measures of schizotypy and psychosis-proneness, although some measures (e.g., the CAPE) primarily assess the presence of depressive rather than disorganised symptoms.

While higher scores across all three domains of schizotypy (as opposed to just one or two) are more highly predictive of a later transition to schizophrenia (Grant et al., 2018), we will be predominantly focusing on its positive dimension, due to its relationship with delusions (Andreasen, 1982; Vollema & van den Bosch, 1995). Positive schizotypy is a subcategory in the Community Assessment of Psychic Experiences (CAPE-42: the measure used by Moritz et al., 2012; van Os, Verdoux, & Hanssen, 1999), and there is a similar positive subscale within the Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE; Mason & Claridge, 2006), termed "unusual experiences". Importantly, these subscales capture belief in delusions or delusion-like ideas, from magical ideations such as belief in witchcraft, to ideas specifically indicative of known delusions, such as persecutory ideas and Capgras syndrome. It is this facet of personality that we expect might predict a greater reliance on intuitive reasoning (e.g., via processing fluency), thereby resulting in higher levels of illusory conviction brought about by stimulus repetition. This idiosyncratic, intuitive sense of conviction may in turn be responsible for belief in the kind of unusual or nontypical ideas observed in clinical and subclinical threshold delusional thinking. Therefore, within the framework of the truth effect phenomenon, we expect individuals higher in positive schizotypy to exhibit a more pronounced effect of repetition on truth evaluation.

Individual differences in the truth effect

A handful of studies has investigated the role of individual differences in the magnitude of the truth effect to date, and these have yielded relatively few positive findings (see Dechêne et al., 2010, for a review). Researchers have explored the influence of variables such as age, cognitive style and ability, scepticism, and probable diagnosis of schizophrenia on

susceptibility to the repetition-induced truth effect. The general trend is that the truth effect is largely robust against individual differences, however the possibility of an influence of schizotypy has received some preliminary supportive evidence.

Age

Research supports the idea that a reduction in memory and recollection performance tends to occur with age (Craik, 2000). Further, there is evidence of poorer recognition and source memory contributing to illusory truth in adults (Law et al., 1998; Skurnik et al, 2000). The rationale behind this hypothesis is that a decrease in source recollection of a repeated item might elicit a greater reliance on intuitive reasoning processes – such as fluency effects – in judgements among older adults, resulting in an increased perception of truth for repeated items. The findings of Law et al. (1998) and Skurnik et al. (2000) support this hypothesis; however, because these studies were interested in differences in illusory truth as facilitated by age-related variations in memory, time was a crucial variable in eliciting the effect. Importantly, these authors noted that the length of delay between first and second presentation of an item moderated differences in the repetition-induced truth effect *via its influence on source memory*.

More recently, a study by Parks and Toth (2006) investigated the influence of age on susceptibility to the truth effect and found no significant relationship between the two variables. Importantly, this study presented both encoding and test phases relatively close in time (as in the present study), unlike previous investigations. The authors noted that the impact of age on reliance on processing fluency had only been demonstrated as a result of memory effects, and argued that this did not indicate a direct relationship between age and the truth effect. In contrast to earlier research, Parks and Toth (2006) specifically investigated age-related variations in the truth effect as facilitated by manipulations in processing fluency only, finding that no differences emerged between younger and older adults. The authors thus concluded that

age does not appear to impact the magnitude of the truth effect via a direct influence on processing fluency.

Put simply, it appears that age-related differences in the truth effect appear to be driven primarily by the length of delay between first and second presentation of an item, with longer delays of one day or more essential for eliciting a larger truth effect among older adults – presumably via age-related differences in source memory. In contrast, no age-related differences in the truth effect were observed during shorter delays when differential source memory was less likely to impact performance. This pattern implies no role for age in modulating processing-fluency, which is considered to be the primary and most direct metacognitive factor driving the truth effect (Reber & Schwarz, 1999; Unkelbach, 2007).

Cognitive Ability, Need for Cognitive Closure, and Cognitive Style

Interestingly, some research has found the repetition-induced truth effect to be robust against cognitive ability, cognitive thinking style, and need for cognitive closure (De keersmaecker et al, 2020). De keersmaecker et al. (2020) reasoned that individual differences in cognitive ability and style – such as analytic versus intuitive thinking – might differentially affect reliance on processing fluency and, consequently, on susceptibility to the truth effect. The authors took measures of their key variables and conducted a series of standard truth effect experiments to examine whether any relationships might emerge between cognition and illusory truth. Delay between first and second presentation of trivia statements ranged from minutes (i.e., within the same experimental session) to 7 days, which controlled for any confounding influence of delay length on the truth effect. Cognitive ability was measured using a shortened version of the Wilde Intelligence Test (Kersting, Althoff, & Jager, 2008), and the Ammons Quick Test (QT; Ammons & Ammons, 1962). Need for cognitive closure – the preference for having an answer (*any* answer) to a given question over confusion or ambiguity

(NFC; Kruglanski, 1990) – was measured using the NFC scale (Roets & Van Hiel, 2011). Cognitive style (i.e., analytic versus intuitive thinking and decision-making style) was determined using the Cognitive Reflection Test (CRT; Frederick, 2005) and an additional self-report measure. The authors compared differences in the magnitude of the truth effect across these three variables over seven different experiments. Across all these variants, no significant relationships emerged between any of the individual difference measures and susceptibility to the truth effect.

Surprisingly, the authors found no effect of intuitive thinking style on truth; however, this does not necessarily indicate there will be no difference in the present investigation. Our analysis focuses more on involuntary, nonconscious use of intuitive processes, rather than a preference for an intuitive 'thinking style' as indicated by a self-report measure. Interestingly, there is some positive evidence in adjacent domains. For example, Newman et al. (2020) found evidence that, under some conditions, need for cognitive elaboration influenced performance on truth assessment tasks that were sensitive to processing fluency effects (see also Barchetti et al, 2022, for effects of "bullshit receptivity"). Thus, it appears that there is still a gap in the current literature, where consensus regarding illusory truth and differences in cognition and personality has not yet been reached. Finding those areas where differences do exist will be crucial in understanding how this phenomenon operates within our population, and for identifying which individuals might be more prone to a reliance on processing fluency in judgements of truth.

Schizophrenia

As discussed in previous chapters, a pilot study by Moritz et al. (2012) found evidence to support the hypothesis that a probable diagnosis of schizophrenia – particularly the presence of positive symptoms of schizophrenia – may influence susceptibility to the truth effect. The

authors compared the magnitude of the truth effect for emotional and neutral items, and across the three participant groups (i.e., nonclinical, clinical with low positive symptoms, clinical with high positive symptoms). A significant truth effect was observed, and the size of the truth effect for delusion-relevant items was significantly larger for patients high in positive symptoms, than for those in the low symptom and nonclinical groups. Further, the size of the truth effect for delusion-relevant items was positively correlated with positive scores in the CAPE, while no relationship was found between the truth effect and negative or depressive symptom scores. Finally, the truth effect for neutral items did not share a significant relationship with scores in any of the subscales in the CAPE. The global pattern of findings is suggestive of a role for delusion-proneness in determining the magnitude of the experienced truth effect.

However, Moritz et al.'s (2012) study was the first to address this hypothesis, and consequently some questions remain unanswered. Furthermore, the study contained some design features that, with the benefit of reflection, could be altered to afford stronger inferential conclusions. Our investigation sought to build upon and refine Moritz et al.'s original finding, yet our results thus far have been unable to replicate the effect of stimulus content (i.e., delusion-relevant versus neutral) on the truth effect. This chapter examines whether schizotypy and delusion proneness was associated with truth effect magnitude across our previous experiments, and if so, whether content type also played a role. The present analysis has the benefit of a more tightly controlled stimulus set than that used by Moritz et al. (2012), and a much larger sample size (n = 302) of pooled data. It is anticipated that the greater experimental power afforded by these amendments will provide a more definitive answer as to the role of individual differences on the truth effect.

Secondary to this question, the present data will allow us to examine whether people high in schizotypy or delusion-proneness are more likely to endorse delusion-relevant content.

There is some evidence to suggest that schizotypy is related to belief in unusual or conspiracy-based ideas (Denovan, 2020). Therefore, we will also explore whether individual differences in delusion-proneness might have played a role in the unexpected findings of Chapter 2 (Experiments 1 and 2), whereby novel delusion-relevant statements returned significantly higher truth ratings than neutral statements. While familiarity was identified as a significant contributor to the discrepancy in Chapter 3, this factor did not explain all the variance. Thus, the present chapter will examine whether schizotypy further explains this variance. Any role of schizotypy in shaping initial or novel truth ratings is important to consider, because the quantification of the truth effect is dependent upon novel truth ratings. It was therefore vital to carefully identify and control for any effect of schizotypy on initial truth ratings in our analyses, so as to isolate the relationship between schizotypy and the truth effect.

Aim and hypotheses

As we have discussed in previous chapters, the experience of processing fluency is used as a form of intuitive reasoning influencing a range of judgements, including trustworthiness, liking, and truth. This chapter asks whether people higher in schizotypal personality traits or delusion-proneness are more prone to rely on processing fluency in judgements of truth, by analysing data collected from Experiments 1, 2 and 5. More specifically, we will be asking: are traits relating to positive symptoms of psychosis or delusion-proneness more indicative of a reliance on processing fluency than other aspects of schizotypy? We will be using scores in the CAPE to see whether the results of Moritz et al. (2012) are replicated here; as well as including scores from the O-LIFE, which taps into both similar and dissimilar facets of schizotypy (e.g., disorganised symptoms).

The aim of this chapter is to explore individual differences in susceptibility to the truth effect in a healthy population. Specifically, it will investigate whether higher scores in trait

schizotypy (particularly positive symptoms) predicts higher evaluations of truth for repeated statements in a truth effect experiment. Secondly, we will explore whether delusion-relevant content elicits a larger truth effect than neutral content among those who are considered higher in positive schizotypy and delusion-proneness. Finally, we will investigate whether individual differences across these traits played a role in the differences observed in novel ratings between statement types in previous chapters – for example, whether schizotypy predicts initial truth ratings for delusion-relevant items.

Our hypotheses are as follows:

(a) There will be a significant positive relationship observed between scores in positive schizotypy and delusion-proneness, and the size of the truth effect in combined data from Experiments 1, 2 and 5.

(b) The relationship between positive schizotypy/delusion-proneness and the truth effect will be greater among delusion-relevant statements than neutral statements; and

(c) Positive schizotypy and delusion-proneness will explain the differences between novel neutral and delusion-relevant item responses in evaluations of truth across Experiments 1 and 2.

The relationship between the truth effect and content type

Please note that, while a global overreliance on intuitive reasoning might predict a larger truth effect across all items, we are also expecting a difference in the truth effect across content type. This is because we predict that, for those more prone to delusions, certain content (i.e., delusion-relevant claims) might elicit a greater reliance on fluency, or a greater experience of fluency, than other content (i.e., neutral claims).

The basis for predicting a stronger truth effect for delusion-relevant material is primarily attributable to the observable commonality across delusional themes, and further strengthened by the preliminary findings of Moritz et al (2012). However, there is some research which also provides some conceptual support for this prediction. Unkelbach and Rom's (2017) referential theory of the truth effect, proposes that the perceived truth of a statement is partly informed by corresponding references in an individual's memory; references which help inform that individual of the meaning of the statement in question (see Chapter 1 for a detailed description of this theory). Given the high volume of fake news, conspiracy theories and startling health claims circulating the internet and the external world, it is possible that claims with 'delusion-relevance' might be more likely to hit upon a mental reference. That is, even if one has not heard any of the delusion-relevant claims presented across these experiments (indeed, we have been careful to reduce familiarity as much as possible), they may have heard others that similarly incriminate the government in large-scale conspiracies, for example. However, for neutral content, the level of internal references may be lower – for example, a reference to the third largest city in Sweden, or information relating to a giraffe's vocal cords. According to the referential theory, every individual has a multitude of corresponding memories and references to measure against externally experienced statements to inform judgements. For those statement which find even partial matches to internal references, the potential for such statements to be judged as true will be higher.

Method

Participants

This study utilised data from Experiments 1, 2 and 5, which were approved by the Flinders University Human Research Ethics Committee (ID: 2671). Participants were recruited

online via Prolific Academic online recruitment platform (https://prolificacademic.co.uk/). Only subjects who reported English as their first language were recruited, and participants received a payment for their time (between 3.13 and 4.20GBP).

To determine whether this number of participants would provide sufficient power to reduce the chances of Type II error, a post-hoc power analysis was conducted using G*Power (version 3.1; Faul et al., 2009). It was determined that a total sample size of 302 would provide sufficient power at the level of .99, for detecting a medium effect size of $f^2 = .15$ in a linear multiple regression analysis with 8 predictor variables. The experimental survey was created and distributed using Qualtrics survey engine (https://www.qualtrics.com), whereby participants were able to access the survey from their personal device. Qualtrics formats surveys to be easily undertaken on a computer, tablet, or mobile phone.

Measures

As outlined in previous chapters, all participants across each experiment completed two measures of schizotypy or psychosis-proneness. These included the short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE; Mason, Claridge, & Jackson, 1995), and the Community Assessment of Psychic Experiences (CAPE-42; van Os, Verdoux, & Hanssen, 1999).

Design

This study used linear regression analyses to determine the influence of each schizotypy and delusion-proneness subscale in the O-LIFE and CAPE on our key dependent variables: the truth effect and average truth ratings. The subscales were unusual experiences, cognitive disorganisation, introvertive anhedonia, and impulsive nonconformity (in the O-LIFE); and positive, negative, and depressive symptom frequency (in the CAPE). Scores for each subscale

were determined for each participant by averaging responses across items related to each subcategory. The truth effect used in this chapter is the between-items truth effect, as a within-items analysis was not possible in Experiments 2 and 5. This means the truth effect was a calculation of difference scores between truth ratings for novel and repeated items in the second phase of each experiment (i.e., truth ratings for novel items in phase 2 were subtracted from ratings for repeated items). These scores were then pooled across the three experiments.

Results

Demographics

Demographic data from 302 participants (158 male) were obtained from Prolific. Mean age of respondents was 34.14 years (*SD* = 12.26), and the majority (85%) were based in the United Kingdom, USA, Canada, New Zealand, and Australia. All were listed as having English as a first language.

Delusion-Proneness and Schizotypy Measures

As psychotic symptoms and trait schizotypy vary greatly across the general population, outliers in the CAPE and O-LIFE scales were not removed from the present analyses (Jaya et al., 2021).

Short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE). Using pooled data from Experiments 1, 2, and 5, scores over most categories resembled previously documented population norms (Mason, Linney, & Claridge, 2005; Fonseca-Pedrero et al., 2015). For each subscale, we inferentially compared the obtained scores of our sample against the standardization sample using z-scores and found no significant differences between our sample and other population norms. The average score for unusual experiences was 4.24 (*SD* = 3.04), with a *z*-score of 0.30, $p = 0.38$. For cognitive disorganisation, the mean was 5.64 (*SD*

= 3.17), *z*-score = 0.42, *p* = 0.34. Average scores for introvertive anhedonia and impulsive nonconformity were 3.71 (*SD* = 2.44), *z*-score = 0.62, *p* = 0.27, and 3.19 (*SD* = 2.16), *z*-score = 0.29, *p* = 0.38, respectively.

Community Assessment of Psychic Experiences (CAPE-42). The structure of the CAPE measure means that not all items are revealed to all users. If participants rate an item as not occurring, they are not asked about the distress elicited. This affords a degree of freedom in scoring the tool. We followed Moritz et al (2012) and scored the distress of non-occurring items as zero. The average weighted score for frequency of positive symptoms was 1.58 (*SD* = 0.43), while the weighted average for negative symptom frequency was 1.96 (*SD* = 0.52), and the average for depressive symptom frequency was 1.97 (*SD* = 0.54). Z-scores were again computed to compare our population means with the healthy population CAPE-42 score averages reported by Moritz et al. (2012). This revealed that the weighted frequency scores from each population are comparable across negative (*z*-score = 0.56, *p* = 0.29) and depressive (*z*-score = 0.41, *p* = 0.34) symptom frequency. However, positive symptom frequency in our sample differed significantly from that of Moritz et al. (2012; *z*-score = 2.06, *p* = 0.02). This means that the average positive symptom score in our sample was 2.06 standard deviations above the average score in Moritz et al.'s sample, indicating that our population experienced a significantly higher frequency of positive symptoms as measured in the CAPE. This is interesting because our sample was unselected, whereas Moritz et al. sought people with a self-identified psychotic disorder. However, when comparing positive symptom frequency in our study with those reported in more recent studies, no significant differences emerged (Moritz et al., 2015: *z*-score = 0.49, *p* = 0.69; Moritz et al., 2017: *z*-score = 1.50, *p* = 0.07). In summary, the present sample was, if anything, higher on the positive delusion-proneness dimension than

Moritz et al (2012) but when viewed more broadly the present sample was not particularly unusual.

The truth effect

To address our first hypothesis, a series of hierarchical multiple regression analyses were conducted to determine the effect of schizotypy on truth ratings for repeated items after controlling for novel truth ratings. It was important to control for novel truth ratings as truth effect scores are calculated by subtracting truth ratings for novel items from truth ratings for repeated items. However, it must be acknowledged that if one group of people (e.g., those high in positive schizotypy) had particularly high ratings for novel items, then there would be a scale compression for this group. That is, in the absence of any additional effect, those high in positive schizotypy would necessarily have lower truth effect scores because a larger value had been subtracted from their ratings for repeated items. While such an effect would bias measurement in the opposite direction to the present hypotheses, it was nevertheless necessary in order to directly compare the truth effect across individual differences.

This was achieved using hierarchical regression, whereby novel truth ratings were controlled by entering them in Step 1 of each analysis. By using average truth scores for repeated items as the independent variable, and including novel truth ratings in Step 1, we were able to determine the role of schizotypy in predicting the magnitude of the truth effect. All remaining predictors, being each subscale from the O-LIFE and CAPE, were entered in Step 2. Importantly, the positive symptom frequency and unusual experiences subscales reflect our delusion-proneness and positive schizotypy variables and are therefore our key predictors of interest here.

Table 12

Hierarchical regression output for schizotypy subscale variables (Total Truth Effect)

	t	*b*	*SE*	*p*
1. Positive symptom frequency	.44	.06	.13	.66
2. Negative symptom frequency	.05	-.01	.10	.96
3. Depressive symptom frequency	1.43	.14	.10	.16
4. Unusual experiences	2.33	.04	.02	*.02
5. Cognitive disorganisation	.04	<.001	.02	.97
6. Introvertive anhedonia	.98	-.02	.02	.33
7. Impulsive nonconformity	2.50	-.05	.02	*.01
8. All novel truth ratings	6.37	-.53	.08	*<.001

** Denotes p values indicating significance*

In Step 1, novel truth ratings explained a significant 11.2% of the variance in repeated item truth ratings, $R^2 = .112$, $F(1,300) = 37.99$, $p < .001$. In Step 2, trait schizotypy explained a significant, additional 4.9% of the variance, $R^2_{Change} = .049$, $F_{Change}(7,293) = 2.43$, $p = .02$. Overall, schizotypy and novel truth ratings explained a total of 16.1% of the variance in repeated item truth ratings, $R^2 = .161$, $F(8,293) = 7.03$, $p < .001$. Regression output for predictor variables are depicted in Table 12.

The analysis provides some support for the hypothesis that positive schizotypy predicts the magnitude of the truth effect. The significant predictors emerging from this analysis, after controlling for novel truth, include unusual experiences and impulsive nonconformity. The measured effect size of the relationship between unusual experiences and the truth effect ($b = .04$) indicates that with every one-point increase in scores for unusual experiences (0 – 12 categorical range), there will be an average increase of .04 points in the magnitude of the truth effect. Unexpectedly, positive symptom frequency did not emerge as a significant predictor in

the present analysis, however; impulsive nonconformity shared a significant negative relationship with truth ratings for repeated items.

Delusion-relevant vs neutral items. In addressing our second hypothesis, two hierarchical multiple regressions were performed to compare how schizotypy and delusion-proneness predict the truth effect across delusion-relevant items and, separately, neutral content types. Once again, novel truth ratings were controlled for in Step 1, while all other schizotypy variables were entered in Step 2. Tables 13 and 14 depict the variable outcomes of these analyses.

Table 13

Hierarchical regression output for schizotypy subscale variables (Delusion-Relevant Truth Effect)

	t	*b*	*SE*	*p*
1. Positive symptom frequency	.83	.12	.14	.40
2. Negative symptom frequency	1.37	-.15	.11	.17
3. Depressive symptom frequency	1.50	.16	.11	.14
4. Unusual experiences	2.02	.03	.02	*.04
5. Cognitive disorganisation	.63	.01	.02	.53
6. Introvertive anhedonia	.89	-.02	.02	.38
7. Impulsive nonconformity	2.30	-.05	.02	*.02
8. DR novel truth ratings	6.22	-.46	.07	*<.001

** Denotes p values indicating significance*

When looking at the delusion-relevant items only, schizotypy and novel truth ratings explained a significant total of 23.3% of the variance in truth ratings for repeated items, R^2 = .233, $F(8,293) = 11.15$, $p < .001$. In Step 1, novel delusion-relevant truth ratings explained a significant 19.1% of the variance in repeated delusion-relevant item truth ratings, R^2 = .191,

$F(1,300) = 70.89$, $p < .001$. In Step 2, trait schizotypy explained a significant, additional 4.2% of the variance, $R^2_{Change} = .05$, $F_{Change}(7,293) = 2.43$, $p = .03$. Once again, unusual experiences and impulsive nonconformity emerged as significant predictors of the truth effect, with unusual experiences sharing a positive relationship and impulsive nonconformity a negative relationship with truth ratings for repeated, delusion-relevant statements.

Table 14

Hierarchical regression output for schizotypy subscale variables (Neutral Truth Effect)

	t	*b*	*SE*	*p*
1. Positive symptom frequency	.08	-.01	.16	.94
2. Negative symptom frequency	1.20	.15	.12	.23
3. Depressive symptom frequency	1.04	.12	.12	.30
4. Unusual experiences	2.01	.04	.02	.05
5. Cognitive disorganisation	.46	-.01	.02	.65
6. Introvertive anhedonia	.84	-.02	.02	.40
7. Impulsive nonconformity	2.08	-.05	.02	*.04
8. Novel truth ratings	8.33	-.68	.08	*<.001

** Denotes p values indicating significance*

When looking at neutral items only, our predictors once again explained a significant 29.9% of the variance in truth ratings for repeated items, $R^2 = .299$, $F(8,293) = 3.60$, $p < .001$. In Step 1, novel neutral truth ratings explained a significant 21.7% of the variance in repeated neutral item truth ratings, $R^2 = .217$, $F(1,300) = 14.76$, $p < .001$. In Step 2, however, trait schizotypy did not explain any additional variance in truth ratings for repeated, neutral items, $R^2_{Change} = .04$, $F_{Change}(7,293) = 1.95$, $p = .06$.

While the effect sizes for schizotypal traits in the above analyses are relatively small, there does appear to be a statistical difference in the role of schizotypy on the truth effect

depending on statement type. While unusual experiences and impulsive nonconformity both significantly predict truth for repeated delusion-relevant items, this effect does not hold for neutral items. Thus, there is some evidence to support our second hypothesis.

Truth Ratings

Our third hypothesis predicted that positive schizotypy and/or delusion-proneness scores would explain the differences in baseline ratings between neutral and delusion-relevant items in previous chapters. First, a paired samples t-test was conducted to determine whether the data from Experiments 1, 2 and 5 still returned a significant difference in baseline (i.e., novel) ratings between neutral and delusion-relevant content once pooled together. This revealed a significant difference did exist across the data, with novel delusion-relevant items receiving higher truth ratings than neutral items, $t(301) = 2.69$, $p = .008$, $d = 0.15$.

To better understand this pattern of results, two multiple regression analyses were conducted. This approach allowed us to determine whether schizotypy was related to novel ratings for neutral and/or delusion-relevant items through analysing each statement condition separately. In the first analysis, novel delusion-relevant truth ratings were included as the dependent variable, with each subscale of the O-LIFE and CAPE included as predictor variables. In the second analysis, novel neutral truth ratings were the dependant variable. The results are illustrated below in Tables 15 and 16, respectively.

Table 15

Multiple regression output for schizotypy subscale variables (Delusion-Relevant Novel Truth)

	t	*b*	*SE*	*p*
1. Positive symptom frequency	2.15	.24	.11	*.03
2. Negative symptom frequency	.51	-.04	.09	.61
3. Depressive symptom frequency	1.30	-.11	.08	.20
4. Unusual experiences	2.80	.04	.01	*.005
5. Cognitive disorganisation	.28	-.004	.01	.78
6. Introvertive anhedonia	.86	.01	.01	.39
7. Impulsive nonconformity	.09	-.001	.02	.93

** Denotes p values indicating significance*

When looking at delusion-relevant item ratings, schizotypy explained a significant total of 9.5% of the variance in novel truth, $R^2 = .095$, $F(7,294) = 4.41$, $p < .001$. Importantly, only our positive schizotypy scales – unusual experiences and positive symptom frequency – emerged as significant predictors of the variance in novel ratings across this item condition. This finding is in support of our initial hypothesis.

Table 16

Multiple regression output for schizotypy subscale variables (Neutral Novel Truth)

	t	*b*	*SE*	*p*
1. Positive symptom frequency	.35	.04	.11	.73
2. Negative symptom frequency	.08	.01	.09	.94
3. Depressive symptom frequency	1.11	-.09	.08	.27
4. Unusual experiences	.05	<.001	.01	.96
5. Cognitive disorganisation	.36	.01	.01	.72
6. Introvertive anhedonia	.73	.01	.01	.46
7. Impulsive nonconformity	.97	.02	.02	.33

In contrast, when looking at neutral items, schizotypy was no longer a significant predictor of the variance in novel truth ratings, $R^2 = .010$, $F(7,294) = .42$, $p = .89$. These findings suggest that positive schizotypy influenced initial perceptions of truth among delusion-relevant – but not neutral – statements across Experiments 1, 2 and 5.

Discussion

The findings of the present chapter suggest that positive schizotypy selectively influences belief, presumably by an increased experience of processing fluency when performing judgements of truth. Specifically, the degree to which a person reports unusual experiences predicts the size of the repetition truth effect, and this relationship only held when the belief being assessed was thematically consistent with delusional beliefs. Additionally, it appears that positive schizotypy also predicts truth ratings for novel delusion-relevant content, prior to processing fluency manipulations. The potential explanations and implications of these findings are discussed below.

The truth effect

Unusual Experiences. As indicated by our regression analysis, there was a significant positive relationship observed between the truth effect and unusual experiences in the O-LIFE, which captures the positive dimension of schizotypy. As previously mentioned, positive schizotypy is exemplified by the presence of unusual sensory, perceptual, and cognitive experiences, and is related to positive symptoms of psychosis such as delusions and hallucinations (Nelson et al., 2013). After controlling for novel truth ratings, scores in unusual experiences predicted the magnitude of repeated truth ratings across all items, with higher scores predicting a larger truth effect. However, once ratings for neutral and delusion-relevant items were analysed separately, it was revealed that this effect only held for items with

delusion-relevant content (it was non-significant with a marginal $p = .06$ for neutral items). While the effect of unusual experiences on the truth effect was generally small, it was robust across analyses and controls, and it appeared relatively specific to the positive dimension of schizotypy. This supports our first hypothesis: that positive schizotypy would predict the truth effect. Further, they provide support for our second hypothesis, being that this effect would be primarily evident among delusion-relevant item content.

The results of this study provide some evidence for a subtle, but detectable, increased reliance on processing fluency, and thus greater susceptibility to implicit, unconscious influences on truth judgements, among individuals who are higher in positive schizotypy and delusion-proneness. Further, it appears this effect might be at least partly dependent upon the thematic content of the information at hand. That is, a greater reliance on intuitive cues for truth (such as processing fluency) occurs in response to content related to themes of delusional beliefs, such as conspiracy, persecution, and contamination. The evidence for this phenomenon is small but significant, as might be expected when assessing a (presumably) largely non-clinical cohort. On the assumption of schizotypy existing on a continuum with clinical psychosis, any plausible influence of schizotypy on illusory or inappropriate belief conviction would be expected to be subtle in individuals who do not present with marked clinical delusions. Interestingly, this relationship held exclusively for the positive dimension of schizotypy within the O-LIFE measure, rather than its equivalent dimension in the CAPE.

Positive Symptom Frequency. Unexpectedly, there was no relationship observed between positive symptom frequency in the CAPE and the truth effect across experiments. Essentially, this indicates that positive symptom frequency does not predict the magnitude of the truth effect, and therefore is not related to a greater reliance upon processing fluency in judgements of truth.

Not only do these results contradict our original hypothesis, but they also run counter to the findings of Moritz et al. (2012). Moritz et al. reported a significant relationship between CAPE positive symptom scores and the truth effect. The tenfold larger sample size of the present analysis (n = 302 repeated measures design versus approximately n = 30 per group in Moritz et al., 2012) represents an obvious difference between the studies. Viewed in isolation, this suggests that the present study merely supersedes the earlier one due to its substantial increase in experimental power.

However, an exclusive focus on sample size misses other potentially important differences between the studies, which allow for a more nuanced and complementary interpretation. For example, another key difference is that our sample did not include a probable clinical population, while Moritz et al. included both a self-reported clinical and nonclinical population to make comparisons in the magnitude of the truth effect. It would therefore be tempting to attribute any difference in findings to the fact that our sample demonstrated fewer symptoms, or were lower in delusion-proneness, than the probable clinical population measured by Moritz et al. However, if anything, the reverse pattern was true. Our sample scored significantly higher in positive symptom frequency in comparison to the healthy population scores in Moritz et al. (M = 1.58 (0.43) and M = 1.21 (0.18), respectively), as reflected by the significant z-test (p = 0.02). Even more compelling is the finding that our overall population scores were actually more comparable to those of the clinical sub-group in Moritz et al.'s study (z-score = 0.73, p = 0.77). This suggests that our sample populations did not differ significantly in psychosis-proneness and strengthens the findings of the present analysis. Further detailed analysis of the differences between this study and Moritz et al., and the implications this holds for future research in the area, are held over to the General Discussion.

Positive Symptom Frequency vs Unusual Experiences

The absence of a relationship between the truth effect and the CAPE is even more surprising given that unusual experiences in the O-LIFE – our other measure of positive schizotypy – *did* in fact predict the magnitude of the truth effect. The first possible explanation for this discrepancy lies in the level of correlation between the two positive schizotypy subscales ($r = .64$, $p < .001$). Of course, correlations between different measures of schizotypy are routinely observed (Mason, 2015), and indeed support the validity of the claim of a unitary psychosis continuum (van Os et al., 2000). However, in this specific case in which both CAPE and O-LIFE were entered as predictors in the same regression analysis, there is the risk that their relationship with each other detrimentally impacted the validity of the results drawn from the analysis. The general term for this concern is *multicollinearity effects* (Alin, 2010).

Fortunately, there are guidelines and tools to minimize or avoid this concern. A multicollinearity check revealed that the relationship between these two subscales did not significantly impact our results. Using variance inflation factors (VIF) with the very low cut-off of 4 as an indicator of multicollinearity (Fox, 2016), none of our predictors emerged as problematic in our regression. In fact, no predictor exceeded a VIF of 2.45, with unusual experiences and positive symptom frequency returning values of 1.99 and 2.10, respectively. Given that the most common VIF cut-off scores used are 5 and 10 (Fox, 2016), we can be confident in concluding that the present findings have not been affected by multicollinearity. Therefore, it is unlikely that the presence of unusual experiences effected the variance output calculated for positive symptom frequency in the present analysis.

Scale differences in measuring schizotypy

Another possible explanation for our results lies not in the similarities, but in the *differences* between our two measures. Some of these differences include the way each scale

was composed, what each claim to measure specifically, and the purpose intended behind their development. Due to these differences, one might argue that the CAPE is a more conservative measure, developed to identify particularly high levels of schizotypy (and thereby indicate clinical risk), as distinct from the O-LIFE, which is designed to be sensitive to small variations in the underlying personality construct between individuals (quite distinct from their clinical risk profile). This might explain the discrepancy between the two scales in their relationship with the truth effect; that is, while they measure the same construct, they measure the construct differently and to differing degrees. However, a detailed consideration of this hypothesis, and its implication for measuring any association with the truth effect, is reserved for the final chapter.

Schizotypy and novel ratings of truth

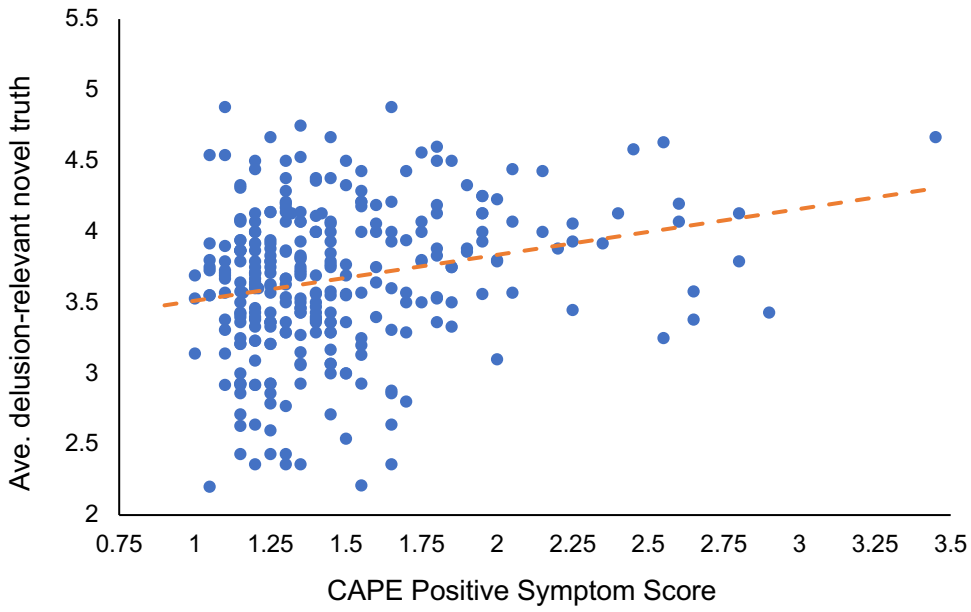
Despite the difference between the O-LIFE and CAPE in their capacity to predict the truth effect, positive schizotypy as captured by both scales significantly predicted novel truth ratings for delusion-relevant items. That is, individuals who scored higher in positive schizotypy rated novel delusion-relevant (but not neutral) items more true upon first encounter than those who were lower in positive schizotypy, showing a small-to-medium effect (see Table 15 and Figure 13). This finding is consistent with previous evidence supporting a relationship between positive schizotypy, conspiracy beliefs, and intuitive-experiential thinking style (Denovan et al., 2020).

It is notable that only positive schizotypy, but not negative, depressive, or disorganised features, were positively related to endorsement for novel, delusion-relevant content. This provides some support for the idea that perceptual abnormalities, delusion-proneness, or psychosis-like experiences might be related to truth in the context of delusion-like or emotive content. Further, the selectivity of this relationship argues against a potential third variable

explanation, whereby people who are simply more likely to endorse items of any kind (that is, a positive global response bias) score highly on both schizotypy and truth ratings. Such an explanation does not hold, because the effect was selective, and was not present for negative or disorganized schizotypy items, nor for neutral control items on the truth effect task.

Figure 13

*Scatterplot showing the relationship between scores in positive symptom frequency as measured in the CAPE (*x*-axis) and average novel truth ratings for delusion-relevant items (*y *axis;* $p < .001$, $r = .26$).*



Surprisingly, while processing fluency was not experimentally manipulated for novel items, it is evident that statements with delusion-like content felt more true to those higher in positive schizotypy. Perhaps differences in processing fluency or other intuitive, metacognitive processes exist independently of the present manipulation of repetition among this population. Indeed, in the study by Denovan et al. (2020), it was found that positive schizotypy predicted intuitive-experiential thinking, and that preferential thinking style (intuitive vs rational) mediated the relationship between schizotypy and conspiracy beliefs. Therefore, perhaps some

belief themes, including conspiracy beliefs, are preferentially supported by intuitive reasoning rather than by solid, objective evidence, particularly among those with positive trait schizotypy.

It is important here to emphasise that we sought to minimize the influence of pre-existing biases for statements by making sure that half of the statements from each category were true, and half were false. Therefore, complete endorsement or denial of all items was wrong for any given statement type. Yet statements were intentionally chosen to represent facts or claims whose truth value was not widely known, in order to encourage intuitive judgment. Thus, for most people, the majority of responses would have to be guessed, and so people would be particularly disposed to the use of pre-existing values or biases (e.g., intuitive reasoning strategies) to form their initial impressions. If schizotypy more reliably predicts the use of intuitive thinking styles, and belief in unusual ideas such as conspiracy theories, this likely led to the observed relationship between schizotypy scores (on both O-LIFE and CAPE) and subjective endorsement of delusion-relevant ideas.

Impulsive Nonconformity

Unexpectedly, our analyses indicated that impulsive nonconformity significantly predicted the size of the truth effect. The results indicated that an increase in impulsive nonconformity predicted a *decrease* in the magnitude of the truth effect, with each one-point increase predicting a decrease of .05 in the size of the effect. There has been some research to indicate that a minority of people reliably demonstrate a *negative* truth effect, whereby repeated items are less likely to be endorsed as true (Schnuerch et al., 2021). At present, there is no research to suggest which individuals are more likely to demonstrate this negative effect, but the findings of our analyses point to a personality trait – impulsive nonconformity – that could partially explain the phenomenon.

However, it is important to treat this interpretation with caution, as measured nonconformity could well refer to failing to comply with task instructions, thereby hampering measurement of the truth effect itself. If a person was impulsively non-conformist, one course of action they might take is to cease performing the task meaningfully partway through. The present task featured design elements to protect against biases induced by such non-compliant behaviour, such as measuring novel ratings and repeated ratings at the same time, in randomly interleaved trials. Therefore, any change in state should affect both items equally. However, the protection afforded by these design elements was not complete. If someone adopted a radically non-compliant approach (e.g., rating "1" for every item after item #46), then the measured truth effect for that person would approximate 0. While a pattern of responding as extreme as this would have been noticed and manually screened out in our analyses, it remains possible (perhaps even likely) that people with a tendency for non-compliance may have intermittently ceased performing the task, or may have reduced their effort. Either of these more subtle acts of non-compliance would have gone unnoticed, and both would tend to lead to an underestimate of the truth effect (a generally positive parameter would be reduced towards zero). Such a pattern of responding could lead a negative relationship between non-compliance and measured truth effect, such as the one observed. Hence, this negative valenced effect must be treated cautiously, as selective non-compliance with the task would tend to produce a similar effect.

Chapter 5: The Effect of Referential Statements on Truth

We know that the fluency with which we process a given piece of information – such as an idea or trivia claim – influences the level of truth we assign to it (Reber & Unkelbach, 2010; Reber & Schwarz, 1999). We also know that repetition can directly influence our experience of fluency, thereby increasing judgements of truth for that statement (Dechêne et al., 2010). Chapters 2 and 3 established that delusion-relevant belief content did not increase the magnitude of the truth effect; if anything, delusion-relevant statements supported a smaller truth effect. While we have thoroughly investigated belief thematic content, we have not yet addressed the subject of many delusional beliefs: the self. Perhaps metacognitive processes such as fluency behave differently with self-related content, whereby it is experienced as more fluent or more salient than information relating to others.

While this specific hypothesis remains untested (to our knowledge), it is clear that self-related information is remembered and processed differently to self-irrelevant information (Symons & Johnson, 1997), but that the nature of this differential processing can vary with task parameters (Klein, 2012). There is also basis to hypothesise a stronger truth effect for self-related material based on Unkelbach and Rom's (2017) referential theory of the truth effect, wherein the authors propose that the judged truth of a statement is informed – at least in part – by meaningful corresponding references in one's memory (see Chapter 1 for a detailed description of this theory). Every individual has a multitude of corresponding memories and references to measure against externally experienced statements about the self, thus strengthening the potential for such statements to be judged as more true. This chapter will seek to explore this idea through conducting a truth effect experiment using self-relevant and self-irrelevant (i.e., relating to others) information. We will also compare the truth effect for

these types of information across trait schizotypy and delusion-proneness, to see whether the truth effect for self-relevant content behaves differently across the psychosis continuum.

Psychological disorders and self-related beliefs

We know that beliefs about oneself are very powerful influencers in our lives, shaping our emotions, cognitions, and behaviours (Beck, 2020; Westbrook et al., 2011). This is particularly evident in the clinical world, where erroneous, overvalued, or delusional ideas about oneself can be indicative of a range of psychiatric illnesses, including mood disorders (Beck, 1979), eating disorders (Jones et al., 2007), and psychotic disorders such as schizophrenia (American Psychiatric Association, 2013). In depression, for example, core beliefs regarding one's lack of worth are common, impacting on a person's ability to form and maintain social relationships, and to engage in goal-oriented activities (Beck, 1979; Maddux & Meier, 1995). In eating disorders, a dysmorphic view of one's body elicits negative beliefs about oneself, leading to overwhelming shame, guilt, and self-criticism, and motivating unhealthy eating behaviours in order to "fix" the perceived bodily flaws (Garner et al., 1982). In schizophrenia in particular, delusional beliefs such as believing one is being watched, controlled, or persecuted in some way represent a significant barrier to living a functional, healthy life, increasing feelings of isolation, fear, and paranoia. These beliefs and subsequent emotions impact upon one's ability to engage meaningfully in numerous aspects of daily life, affecting social, financial, occupational, and health and treatment outcomes (Freeman et al., 2002).

Across all of these disorders, we notice that overvalued beliefs and delusions are almost universally related to the *self*, rather than to another person or group. Beliefs about one's own worth, capabilities, appearance, or safety, form the basis for painful emotions and dysfunctional behaviours, acting as a barrier to an individual's mental and physical health and wellbeing. In

schizophrenia and other psychotic disorders particularly, many commonly held delusions are self-referential, perhaps most notably in delusions of reference (Startup & Startup, 2005). This type of delusion refers to the belief that neutral, external events – such as television or radio communications – are specifically referring to oneself, and that these messages hold personal significance (Startup & Startup, 2005). Another pertinent example is grandiose delusions, whereby an individual believes he or she has special powers, inflated worth, or superior intelligence than other people (Knowles et al., 2011). Even in persecutory delusions, where erroneous beliefs are held about the threat or danger inherent in another person or group, the belief is most often centred around the threat posed to *oneself* (Freeman et al., 2002). In the clinical world, therefore, it seems that problematic beliefs held with the greatest conviction tend to be self-referential, particularly among delusions. However, we do not yet know *why* such beliefs tend to be self-referential, or further, how self-referential ideas become prone to inappropriately high levels of conviction. Is there something about self-referential ideas that render them more easily manipulated, or susceptible to intuitive reasoning processes, than other forms of information?

Intuitive reasoning and self-relevant information

It is possible that self-referential information is simply more salient in general, and therefore more prone to intuitive reasoning strategies such as cognitive biases and heuristics. Indeed, we know that salient, self-relevant information must be readily perceived and processed in order to make rapid decisions in response to novel or dangerous situations (Corbetta et al., 2008). Further, there is evidence to show that self-relevant information is processed automatically and effortlessly, requiring less attentional resource allocation than self-irrelevant information, or information relating to others (Bargh, 1982; Tacikowski et al., 2017). There is also evidence for an increased susceptibility to cognitive biases – such as

overconfidence – in response to information that holds personal relevance, when compared to self-irrelevant material (Ertac, 2011). As we know that intuitive reasoning strategies are automatic, effortless, and prone to cognitive biases (Kahneman, 2011), it appears then that self-related ideas might be more prone to this type of reasoning. Effectively, it seems that the readiness with which we are open to receiving self-relevant information, and the speed with which we must typically process and make meaning of such information, renders us more susceptible to the use of intuitive reasoning strategies.

If it is true that self-relevant information is more readily evaluated via intuitive reasoning, then this might also extend to an increased reliance on metacognitive processes – such as experiential processing fluency – when making judgements about self-relevant material. This would be reflected by a greater increase in perceptions of truth for self-relevant information that has been repeated (or otherwise manipulated in terms of fluency) than self-irrelevant information. For people who are particularly prone to delusions, this effect might be even more exaggerated than for those who are not delusion prone, potentially due to disruptions in these metacognitive processes. This could help account for the development of inappropriately high conviction for self-referential ideas as observed in clinical delusions in schizophrenia.

There is some research to support the hypothesis that self-relevant material elicits a greater reliance on intuitive reasoning strategies. For instance, there has been some evidence to suggest that the hastiness bias (a cognitive bias similar to the 'jumping-to-conclusions' bias) is more pronounced in response to self-relevant, emotional content. In a study by Dudley et al. (1997), participants were asked to complete one of two probabilistic reasoning tasks, one of which was neutral and one which was emotional (i.e., relevant to the participant). The study population included individuals with depressive symptoms, individuals with delusions, and a

nonclinical control group. As the authors predicted, individuals with delusions displayed a significantly greater hastiness bias than depressive and nonclinical participants across both neutral and self-relevant reasoning tasks, requiring fewer pieces of evidence before making a decision. Perhaps even more interesting, however, was the finding that all three participant groups displayed a more pronounced hastiness bias in response to self-relevant or emotional content. That is, while people with delusions tended to make judgements with less evidence than healthy or depressive controls in general, all participants were susceptible to an increased hastiness bias when making judgements about self-relevant information. These findings provide support for our two central hypotheses: Firstly, that people might be more prone to rely on intuitive strategies – such as heuristics and biases – when evaluating self-related material rather than neutral (or nonpersonal) material; and secondly, that this tendency might be even more pronounced in people with delusions.

Interestingly, a study by Larivière et al. (2017) demonstrated marked differences in functional brain activity associated with self-referential processing between schizophrenia patients (with and without current delusions of reference) and healthy controls. Specifically, the findings revealed hyperactivity in the anterior cortical midline structures (CMS) among those with active delusions, and greater activity in this region was associated with greater delusion intensity across all three groups. There were also functional differences observed within the CMS network between patients with and without active delusions, and between patients and healthy controls. This research suggests that neurocognitive differences exist between people with and without delusions of reference, and that these differences are reflected in the self-referential processing areas of the cortex. While these investigations have not yet been extended to other forms of delusions, it may be possible that individual differences in the

processing of self-relevant information also exist in relation to other types of beliefs, and across the psychosis continuum more broadly.

Aim and hypotheses

This chapter seeks to better understand the role of fluency effects in judgements of truth for self-relevant material compared to statements referring to others. The investigation will be carried out using a standard repetition truth effect experiment, prompting subjective truth ratings for descriptive statements rather than the usual trivia claims used in these experiments. To our knowledge, this is the first investigation of the truth effect measuring evaluation of subjective, descriptive information, in that none of the statements have an objective truth status. Rather, truth judgement is entirely based on participants' personal opinions and experiences, and there is no correct or incorrect answer to any one statement in the experiment. For this reason, the statements used in this study may in fact be more prone to reasoning based on intuition, thereby rendering them more susceptible to manipulations of fluency.

We will also investigate whether individual differences across the psychosis continuum – as measured using self-report schizotypy and psychosis-proneness scales – differentially impact the magnitude of the truth effect for self-referent and other-referent statements. Given that we observed a significant effect of positive schizotypy (in the O-LIFE) on truth for delusion-relevant statements in our previous chapter, we predict that self-related statements will yield a greater truth effect among those higher in positive trait schizotypy than those lower in the trait. Once again, the O-LIFE and CAPE measures will be our method of assessment for schizotypy and delusion-proneness, and we will be primarily concerned with the 'unusual experiences' and 'positive symptom frequency' subscales in each measure, respectively.

Our hypotheses are as follows:

(a) There will be a significant truth effect observed across both conditions, with repetition eliciting an increase in truth ratings for subjective, descriptive statements.

(b) The truth effect will be greater for self-referent than other-referent items.

(c) Positive schizotypy and delusion-proneness as measured by the O-LIFE and the CAPE will predict the magnitude of the truth effect for self-related items.

Method

Participants

This study was approved by the Flinders University Human Research Ethics Committee (ID: 2671). Three hundred participants (125 male) were recruited online via Prolific Academic online recruitment platform (https://prolificacademic.co.uk/). Mean age of respondents was 37.78 years (*SD* = 12.57), and the majority (78%) were based in the United Kingdom, USA, Canada, New Zealand, and Australia. Only subjects who reported English as their first language and had not participated in any previous studies from this project were recruited, and participants received a payment for their time (3.75GBP for approximately 25 minutes).

The experimental survey was created and distributed using Qualtrics survey engine (https://www.qualtrics.com), whereby participants were able to access the survey from their personal device.

The sample size of this experiment was larger than previous studies for two reasons. First, the sample size began as 150, as funding allowed for the collection of a larger sample. Second, a counterbalancing issue was detected following the initial data collection, as the protocol for counterbalancing in this experiment was greatly complicated by the use of descriptive words which varied in valence, length, and emotionality, and which further needed

to be assigned to “I am” or “He/She is” statements across participants. To rectify the issue, it was necessary to run another equal number of participants – that is, a further 150 – to balance the issue in the first scenario.

Materials

Measures of psychosis-proneness. The short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE; Mason et al., 1995) and the Community Assessment of Psychic Experiences (CAPE-42; Stefanis et al., 2002) were again used to assess schizotypal traits and delusion-proneness in the present study. Both the O-LIFE and CAPE are self-report measures which take a relatively short time to complete, and both have been demonstrated as valid and reliable indicators of schizotypy and psychosis-proneness across time (Fonseca-Pedrero et al., 2015; Konings et al., 2006).

Statements. The statements used in the present study were simple, three-word sentences beginning with “I am” or “[Friend’s name] is” and ending with a descriptive word (i.e., adjective) such as intelligent, happy, or disorganised. Development of the list of descriptive words involved accessing a database of 229 frequently used adjectives (YourDictionary; Abbott, 2000), and creating 6 lists of 16 words each, balanced equally in word valence, length (in letters), and syllables (96 words in total). Word valence was determined using a sample of 10 participants, who rated the valence of each word on a 5-point Likert scale ranging from ‘negative’ to ‘positive’. The average ratings for each word were then used to distribute equal numbers of positive and negative words across each of the 6 lists.

Participants were randomly allocated to one of 6 counterbalancing conditions, whereby each list was differentially allocated to repetition and reference types. This careful allocation of words to counterbalanced lists, and subsequent allocation of these lists to different statement types, was to control for any extraneous variables that may have influenced truth ratings (e.g.,

if some words were less well-known, we might not get an accurate reflection of truth for those items, thus impacting the truth effect measured). Presentation of all 96 adjectives was allocated to reference either the self or the other throughout the experiment, with 48 statements referring to each subject. Half the statements (i.e., 48 statements) for each reference type were then repeated in the second half of the experiment.

Design

The effect of repetition on truth ratings was measured using a within-subjects design, in which presentation of self-referent and other-referent statements were manipulated to appear either once or twice throughout the survey. The study had a 2 (repetition: repeated, not repeated) x 2 (statement type: self, other) repeated measures design. The dependent variable was the truth rating response given for each statement, with the truth effect calculated by subtracting the truth rating for repeated statements from the truth rating from non-repeated items. Measures of individual differences in schizotypy were also captured by two questionnaires. Results from these measures were calculated for each participant and separated into different subcategories of schizotypy, and regression analyses were used to determine whether these results predicted truth ratings for delusion-relevant or neutral items, or the size of truth effect itself.

Procedure

Participants were informed that the study aimed to examine how briefly presented words affect how people think, to better understand how and why people form rapid, intuitive judgements about the truth of a statement. After consenting to participate, participants were asked to disclose their age and gender. Participants then completed the O-LIFE and CAPE questionnaires in the same manner as the earlier experiments. Following this, participants were

asked to provide the first name of a close friend, colleague, or family member for the purpose of the study.

Participants then completed the entire truth ratings questionnaire in one sitting, with one statement presented per page, along with a 6-point Likert scale below each statement. These statements were presented as simple "I am…" or "[Friend's name] is…" statements ending in adjectives such as kind, curious, impatient, or analytical, and the truth of these statements was evaluated using a scale ranging from 'Definitely false' to 'Definitely true'. In the first half of the experiment, participants rated 64 descriptive statements, half of which were self-related, and half of which were other-related. Note that people did not rate interestingness in this experiment. While altering the question between phases increased the magnitude of the observed truth effect (see Chapter 2) it was not clear that people could make an interestingness judgment on the present statements (it is unclear how one should assess the phrase "Jennifer is friendly" for its interestingness). Therefore, a single question was used throughout this task.

Directly following this initial set of items, another 64 statements were presented: half of which were repeated from the first set, and half that were novel. The order of presentation of items was randomised within each set (i.e., the initial set was randomized independently of the second set of items). Overall, one third of all items were repeated, totalling 32 statements (half self-related, half other-related), so that each participant rated the truth of a series of 128 statements in total. Three different counterbalancing conditions were created to ensure each of the items was equally likely to be assigned each of the three possible roles: novel item in the initial set, novel item in the secondary set, repeated item (and thus shown in both sets).

The present experiment was designed in such a way that both within-items and between-items truth effect was measured. Calculation of the within-items truth effect was obtained by subtracting response ratings for to-be-repeated items from ratings for repeated

items. Calculation of the between-items truth effect was obtained by subtracting truth ratings for repeated items from ratings for novel items in the second half of the survey.

Results

Delusion-proneness and schizotypy measures

As psychotic symptoms and trait schizotypy vary greatly across the general population, outliers in the CAPE and O-LIFE scales were not removed from the present analyses (Jaya et al., 2021).

Short Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE). Scores over each of the four subscales in the O-LIFE resembled previously documented population norms (Mason & Claridge, 2006; Fonseca-Pedrero et al., 2015). For each subscale, we inferentially compared the obtained scores of our sample against the standardization sample using z-scores and found no significant differences (all $p > .05$) between our sample and other population norms. The average score for unusual experiences was 4.53 ($SD = 2.99$), with a z-score of .40, $p = .34$. For cognitive disorganisation, the mean was 5.22 ($SD = 3.19$), z-score = .61, $p = .39$. Average scores for introvertive anhedonia and impulsive nonconformity were 3.63 ($SD = 2.37$), z-score = .58, $p = .28$, and 2.75 ($SD = 1.97$), z-score = .07, $p = .47$, respectively.

Community Assessment of Psychic Experiences (CAPE-42). For the CAPE, the corresponding frequency response for missing distress response data fields were removed. The average weighted score for frequency of positive symptoms was 1.45 ($SD = .36$), while the weighted average for negative symptom frequency was 1.94 ($SD = .52$), and the average for depressive symptom frequency was 1.96 ($SD = .55$). Z-scores were again computed to compare our population means with the healthy population CAPE-42 score averages reported by Moritz et al. (2012). This revealed that the weighted frequency scores from each population are

comparable across positive, (z-score = 1.33, p = .09) negative (z-score = .51, p = .30) and depressive (z-score = .39, p = .35) symptom frequency.

The truth effect

As previously discussed, the truth effect is calculated by subtracting truth ratings for repeated statements from ratings for novel statements. To determine whether there was a standard truth effect, two separate analyses examined differences in ratings between first and second presentation of the same item (within-items effect) and, separately, between novel and repeated statements in the second half of the experiment (between-items effect). Descriptive statistics are reported in Table 17.

Table 17

Mean Truth Rating (With Standard Deviation) by Statement Type and Repetition Status Across Training and Test Phases

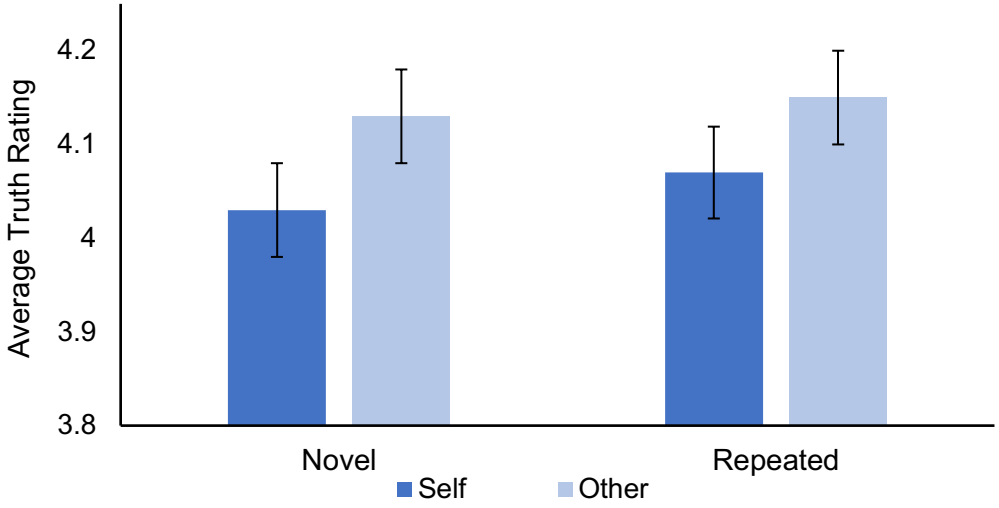
Statement Type	Training		Test	
	Novel	To be repeated	Novel	Repeated
All	4.09 (.41)	4.10 (.41)	4.08 (.41)	4.11 (.44)
Self	4.03 (.50)	4.07 (.49)	4.03 (.50)	4.07 (.52)
Other	4.14 (.50)	4.13 (.51)	4.13 (.50)	4.15 (.54)

Between-items effect. Differences in average truth ratings between novel and repeated items in each statement condition are depicted in Figure 14. To determine whether a truth effect was observed between novel and repeated items in the second half of the experiment, a repeated measures ANOVA with two factors was conducted, including statement type (self vs other) and repetition status (repeated vs novel). Contrary to our predictions, there was no significant main effect of repetition across all statements, meaning that a between-items truth effect was not observed, $F(1,299) = 2.79$, $p = .10$, $\eta^2 = .002$. The main effect of statement type

was significant, indicating a difference in truth ratings between self-referent and other-referent statements, $F(1,299) = 10.61$, $p = .001$, $\eta^2 = .02$. Post-hoc analyses revealed that other-referent ($M = 4.14$, $SD = .45$) statements received significantly higher truth ratings than self-referent ($M = 4.05$, $SD = .43$) statements overall, $t(1,299) = 3.42$, $p < .001$, $d = .20$. Finally, there was no significant interaction observed between statement type and repetition, $F(1,299) = .51$, $p = .45$, $\eta^2 < .001$.

Figure 14

Mean Truth Ratings (With Standard Error Bars) by Statement Type and Repetition Status in the Second Half of the Experiment (i.e., Test Phase)



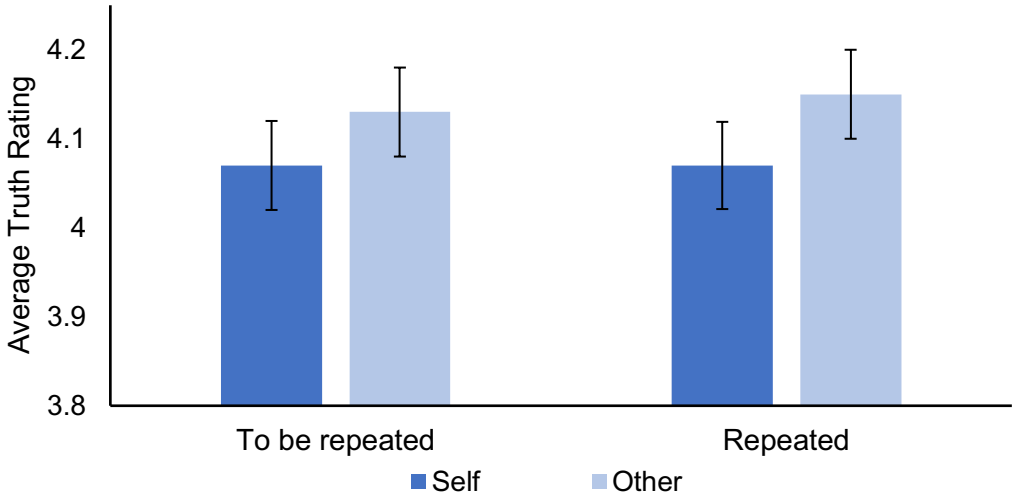
Note. This figure illustrates the difference between truth ratings for novel and repeated items in the experimental test phase.

Within-items effect. A second analysis was carried out to investigate whether a difference in truth ratings occurred between first and second presentation of the same item (i.e., within-items truth effect). A repeated measures ANOVA was used, with two factors: statement type (self vs other) and repetition status (repeated vs to be repeated). The main effect of repetition on truth ratings was not significant, meaning that a within-items truth effect was not

observed, $F(1,299) = 1.93$, $p = .17$, $\eta^2 < .001$. The main effect of statement type was again significant, revealing a difference in truth ratings between self- and other-referent statements, $F(1,299) = 4.22$, $p = .04$, $\eta^2 = .012$. Again, post-hoc comparisons revealed that other-referent ($M = 4.14$, $SD = .51$) statements received significantly higher truth ratings than self-referent ($M = 4.07$, $SD = .49$) statements, $t(1,299) = 2.05$, $p = .04$, $d = .13$. The interaction between statement type and repetition was once again nonsignificant, $F(1,299) = 2.69$, $p = .10$, $\eta^2 < .001$.

Figure 15

Mean Truth Ratings (With Standard Error Bars) by Statement Type and Repetition Status Across Training and Test Phase



Note. This figure illustrates the difference in truth ratings for items before and after repetition.

Schizotypy, referential statements, and the truth effect

As there was no significant truth effect, a full regression analysis of individual differences in the magnitude of the truth effect was withheld. Instead, simple bivariate correlation analyses were conducted with regards to schizotypy and the effect of repetition on truth. No significant associations were observed. For positive symptom frequency in the CAPE, there was no relationship with the truth effect for self-referent ($r = .09$, $p = .13$) or other-referent

(r = -.01, p = .94) items. Similarly, there was no relationship observed between scores in unusual experiences in the O-LIFE, and the truth effect for self-referent (r = .06, p = .30) or other-referent (r = -.01, p = .90) statements.

Schizotypy and novel truth ratings

Putting aside whether the truth effect is modified by schizotypy, a second question pertains to whether self-referential statements are assessed differently in those high in schizotypy to those low in schizotypy. Recall that novel, delusion relevant statements were endorsed at higher rates by high than low schizotypy participants. Two exploratory multiple regression analyses were conducted to test if positive schizotypy was associated with initial ratings of truth amongst self-referent statements, similarly to the preferential endorsement of delusion-relevant content seen in Chapter 4. For contrast, a regression analysis investigated any association between schizotypy and other-relevant statements too. The analysis followed the same format as that used in Chapter 4. Novel truth ratings from the second half of the experiment were entered as the dependent variables, with one analysis containing novel, self-referent truth ratings only, while the other contained novel, other-referent truth ratings.

Table 18

Multiple Regression Output for Schizotypy Subscale Variables (Self-Referential Novel Truth)

	t	*b*	*SE*	*p*
1. Positive symptom frequency	.82	.08	.09	.41
2. Negative symptom frequency	.09	.01	.07	.93
3. Depressive symptom frequency	.28	-.02	.01	.78
4. Unusual experiences	4.57	.05	.01	<.001*
5. Cognitive disorganisation	.37	-.004	.01	.71
6. Introvertive anhedonia	4.58	-.06	.01	<.001*
7. Impulsive nonconformity	.37	-.01	.02	.71

** Denotes p values indicating significance*

When looking at self-referential items, schizotypy explained a significant total of 15.6% of the variance in novel truth, $R^2 = .156$, $F(7,292) = 7.70$, $p < .001$. Importantly, unusual experiences emerged as one of only two predictors of the variance in novel ratings across this item condition. With each increase of 1 point in unusual experiences (a continuous scale ranging from 1-12), average novel truth ratings for self-related items increase by .05. Further, a bivariate correlation analysis revealed that unusual experiences and novel ratings for self-related items were significantly and positively related, $r = .27$, $p < .001$. Introvertive anhedonia also predicted novel truth, but scores in this category shared a *negative* relationship with truth ratings, whereby higher scores predicted *lower* truth ratings for self-referent items. This significant relationship was confirmed in a bivariate correlation, $r = -.24$, $p < .001$.

Table 19

Multiple Regression Output for Schizotypy Subscale Variables (Other-Referential Novel Truth)

	t	*b*	*SE*	*p*
1. Positive symptom frequency	.79	.08	.10	.43
2. Negative symptom frequency	1.26	-.10	.08	.21
3. Depressive symptom frequency	.43	.03	.07	.67
4. Unusual experiences	1.31	.02	.01	.19
5. Cognitive disorganisation	1.92	.02	.01	.06
6. Introvertive anhedonia	4.58	-.06	.01	<.001*
7. Impulsive nonconformity	.09	.002	.02	.93

** Denotes p values indicating significance*

When looking at other-referential items only, schizotypy explained a significant total of 9.3% of the variance in novel truth, $R^2 = .093$, $F(7,292) = 4.27$, $p < .001$. However, only introvertive anhedonia emerged as a significant predictor of truth, again sharing a negative

relationship with truth ratings. This was again confirmed using a bivariate correlation, $r = -.22$, $p < .001$.

These findings indicate that positive schizotypy – as measured by unusual experiences in the O-LIFE – may positively predict novel truth for self-referent, but not for other-referent material.

Discussion

The findings of the current experimental chapter provide novel evidence regarding self-referential content and the truth effect phenomenon. To our knowledge, this is the first study to investigate the truth effect using subjective, descriptive statements. Unfortunately, no overall truth effect was observed, so it was difficult to make clear conclusions regarding whether self-relevant or other-relevant content modified that effect. One possible interpretation is that the truth effect does not occur with descriptive content relating to the self and others, although alternative explanations ought to be considered too.

Familiarity

The truth effect has been shown to occur following repetition of novel content (Hasher et al., 1977). That is, when unfamiliar information is presented and then repeated, people are more likely to judge it true after the second encounter. However, this effect ostensibly depends on content being novel or close to novel. Subsequent repetitions over time have been demonstrated to yield smaller effects – and sometimes no effect – on endorsements of truth (Dechêne et al., 2010; Hassan et al., 2021). Hassan et al. (2021) demonstrated that the greatest increase in perceptions of truth occurs following the second presentation of a statement, with the truth effect becoming smaller with each additional encounter. They concluded that more

frequent repetition does not exponentially increase the truth effect, but rather shares a logarithmic relationship with perceptions of truth.

We observed this effect previously in Experiment 2, where delusion-relevant items received elevated truth ratings when novel, and no further truth effect was elicited following their repetition. In a subsequent item-level analysis we later determined that this was predominantly due to high levels of familiarity for many of the statements, meaning that these items were not truly novel upon first encounter in the experimental procedure (see Chapter 3). This hypothesis was further corroborated when replacement of the items with highest familiarity then yielded a significant truth effect in Experiment 5.

A limitation of the present experiment is that the material being evaluated was not truly novel. The use of frequently used, well-known adjectives, as well as the subject matter being oneself and a close friend, meant that the content was already highly familiar. People are often motivated to reflect on their own personalities (Jenkins & Mitchell, 2011), regardless of the accuracy of such knowledge; any such tendency would attenuate the repetition manipulation used here. Furthermore, people often form close relationships with others based on socially desirable qualities, or characteristics similar to their own (Verbrugge, 1977). Therefore, it is likely that participants may have previously contemplated the extent to which they – and their friend – demonstrate many of the qualities described in this experiment. If the present stimulus statements had been encountered numerous times, perhaps due to spontaneous reflection on one's attributes or those of a close friend, then any effect of a single repetition on judgment may be negligible.

One approach to control for this issue in future investigations would be to compose more complex referential statements for people to evaluate. For example, participants might be presented with unique statements about their likelihood to behave in a particular way in

response to a novel situation (e.g., "I am analytical when faced with a difficult logistics challenge at work"). The participants' evaluation of such a statement would be in terms of whether that behaviour is likely to be true for them in that specific context. This would reduce the likelihood that the specific attribute was familiar at the beginning of the experiment.

An alternative approach is to replace the adjectives used here with unfamiliar descriptive words, or even non-words. While this would only remove familiarity for the descriptive content (adjective) and not the subject of the content (self or other), this approach might allow for less familiar, and therefore previously unconsidered, content to be evaluated.

Word valence

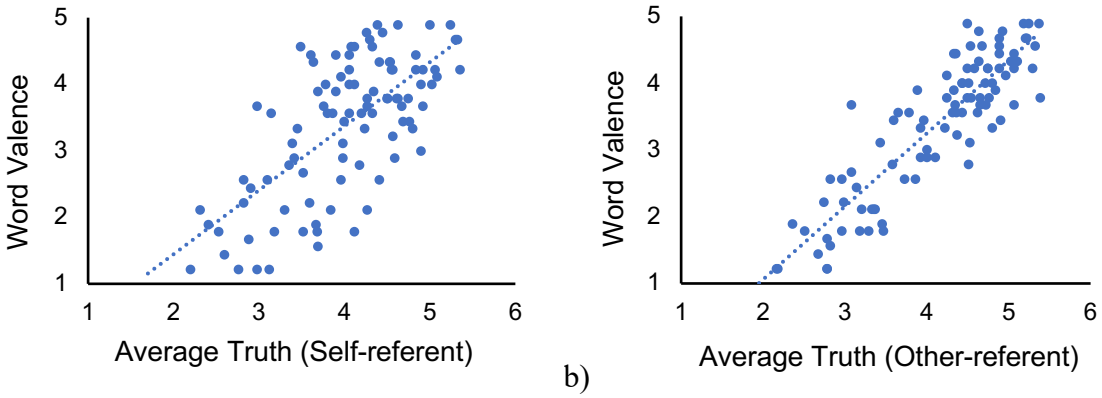
Another potential factor influencing truth in the present study is the valence of the descriptive words used. Word lists in this experiment were carefully controlled to be balanced in terms of average word valency, to ensure that any influence of valence on truth would be evenly distributed across repetition status, reference type, and counterbalancing conditions. Nevertheless, if some truth ratings were more heavily influenced by valence than by repetition status or reference type, this may have significantly affected our results. For example, if novel statements with highly positive words (e.g., "honest") were consistently rated as mostly or definitely true, this could reduce or remove their capacity to return a truth effect following repetition.

An exploratory item-by-item analysis revealed that word valence had a large and significant effect on endorsements of truth for novel items. Specifically, novel truth ratings and valence shared a significant, positive relationship for both self-referent ($r = .68$, $p < .001$) and other-referent ($r = .87$, $p < .001$) statements. This finding reflects a potential confound in our experiment. As can be seen in Figure 16, the influence on valence was medium-to-large, depending on statement type. The items that were more positive in valence yielded higher truth

ratings when novel, which may have caused a ceiling effect for those items, thereby rendering any additive effect of repetition undetectable. Further analysis indicates that 39% of self-referent, and 44% of other-referent items returned novel truth ratings of 5 or 6 (i.e., "mostly true" or "definitely true"). Considering the strong relationship observed between valence and novel truth, it is likely that these highly rated items were most often those containing positive descriptive words.

Figure 16

Scatterplot showing the relationship between word valence and novel truth ratings for (a) Self-referent; and (b) Other-referent statement types.



a) b)

This finding was unexpected but has some empirical evidence to support it. Indeed, there has been previous research suggesting that valence plays an important role in truth judgements (Hilbig, 2009; Jaffè & Greifeneder, 2021). This has been studied in terms of positive versus negative framing, as well as the conceptual valence of information being evaluated – with both demonstrating an influence on truth. For example, Jaffè & Greifeneder (2021) investigated the differential impact of both framing and valence on ratings of truth, and reported that negative framing and negative content valence both independently predicted truth. However, in contrast to the present findings, this research has shown a truth bias for *negative* information across both framing effects and content valence. That is, negative content and

negatively framed content both predicted higher endorsements of truth than positive content and positively framed content. These findings have been demonstrated across several studies, all demonstrating a stronger effect of negative over positive information (Baumeister, 2001; Hilbig, 2009; Hilbig, 2011; Hilbig, 2012; Jaffè & Greifeneder, 2021). In summary, negatively framed statements about external content are more likely to be endorsed as true, while positive statements about ourselves or close friends are more likely to be endorsed as true.

To avoid this issue in future, less variation and strength in word valence might yield lower initial truth ratings, thereby allowing for more accurate observations of the effects of repetition on truth for descriptive, referential statements. This could mean using mostly neutral or negatively valenced words rather than a range of negative and positive words, as well as completely removing highly positive words such as "generous" and "kind" from the experiment.

Memory and consistency

While it has been demonstrated several times that the truth effect does not require a time delay between first and second experimental phases (Unkelbach, 2007; Dechêne et al., 2010), we may have again seen the impact of delay on consistency effects. In Experiment 1 (Chapter 1), no truth effect was observed when people were asked to assess truth across both novel and repetition phases of the experiment. When the first judgment was replaced with an assessment of interestingness in Experiment 2, the same stimuli yielded a large and significant truth effect. It remains possible that the same issue was present in this experiment. People were asked to assess truth twice here because it appeared nonsensical to ask how interesting the self-referential statements were. Yet it appears that intuition was once again overridden by a desire for consistency: 38% of participants in this experiment guessed that the role of repetition was to test memory and consistency across repeated statements. Perhaps a more productive

approach would have been to ask them a different question (e.g., "how pleasant is this statement?") that facilitates repetition, but does not provoke a need for consistency across repetitions. This empirical question remains for future research.

The risk of people seeking to respond consistently would likely have been facilitated by *the self-relevance effect*. The self-relevance effect refers to the phenomenon whereby self-referential information yields better explicit, but not implicit, memory performance (Fujita & Horiuchi, 1998). There are a range of theories as to how this effect occurs (see Klein, 2012, for a review) but a general theme is that the conceptual self is generally well represented in memory, and therefore acts as an excellent foundation for novel information related to that global self-representation. In the context of the truth effect, if people were seeking to maintain consistency in their responding, they would tend to be facilitated in that aim for self-referential statements. Self-referential statements would be more available for recall at the second presentation of that item, thereby facilitating consistent responding. While this benefit would not be present for other-relevant items, there is evidence that statements about others known to the respondent are also subject to memory benefits too (relative to other forms of control conditions; see Symons & Johnson, 1997 for review). In sum, there is reason to believe that the referential stimuli used, in combination with the prompting of truth judgement in both phases of the task, may have reduced the capacity to detect a truth effect in this experiment.

Schizotypy, referential statements, and truth

Contrary to our predictions, trait schizotypy did not appear to influence the truth effect in the present experiment. No subscale across the CAPE or O-LIFE measures predicted the size of the truth effect for self-referent or other-referent statements, but this is difficult to interpret in the absence of a truth effect overall.

In our analysis of trait schizotypy and novel truth, however, we did observe a significant relationship between positive trait schizotypy and self-referential statements. Overall, schizotypy explained 15.6% of the total variance in novel truth ratings for these items. Specifically, unusual experiences in the O-LIFE significantly and positively predicted ratings of truth for novel, self-referent statements, but not for other-referent statements. Positive symptom frequency in the CAPE did not predict novel truth for either statement type.

It is unclear why this particular pattern was observed. We speculate that because self-referential statements were likely not novel at the start of the experiment, then all self-referential statements would have already been subject to repetition – and any effects of repetition – prior to the experiment. Therefore, any differences between individuals in the magnitude of the truth effect could potentially be reflected in differential novel ratings, rather than in the difference between novel and repeated ratings. If so, then the observed relationship between self-referential novel truth ratings and positive schizotypy would suggest the same pattern (i.e., of repetition leading to a greater susceptibility to increased truth) as was evident in the schizotypy-mediated truth effect seen for delusion-relevant items in Chapter 4. However, these ideas are merely speculative, and further investigation of these accounts are best left to future research. Regardless of whether this relationship has anything to do with repetition over time, we do see that self-related material is perceived to be truer with greater degrees of delusion-proneness, suggesting that those who are delusion-prone assess self-relevant material differently to those who are not.

Conclusion

Pre-existing research has consistently shown that the truth effect is robust against previous knowledge and explicit information (e.g., Fazio et al., 2015; Unkelbach & Rom, 2017; Unkelbach & Greifeneder, 2019) Nevertheless, it remains possible that informed opinions can

eliminate the truth effect in the context of judgements about the self and well-known others. The present study was, to our knowledge, the first to explore the truth effect with regards to repeated, referential descriptive statements. While great care was taken to develop a well-controlled, well-powered, and well-balanced experiment, there were some extraneous factors which appear to have persisted despite our experimental controls. Some of these factors include memory and consistency, valence, and familiarity, and it is likely that a combination of these contributed to the null effect observed here. It is possible that future investigations may identify conditions under which referential truth can be influenced by repetition, thus resulting in a truth effect. However, if this is the case, we know that there are limitations to this effect. In summary, we will conclude that repetition does not influence truth for referential statements under the present experimental conditions, suggesting that truth evaluation for simple, descriptive statements is robust against manipulations of processing fluency, and therefore the truth effect itself.

Chapter 6: General Discussion

The purpose of the present thesis was to contribute to current understandings of conviction and its development in clinical delusions. Our key question was how some of the intuitive mechanisms driving perceptions of truth might relate to the development of conviction in delusional beliefs. Specifically, we sought to better understand how reliance upon intuitive evaluation might differ between individuals or between statements of differing thematic content. Intuitive, metacognitive drivers of conviction were modelled using the phenomenon known as the truth effect, whereby repeated statements are processed more fluently and also assessed as more likely to be true than otherwise comparable, novel statements.

The investigation was carried out over four empirical chapters, and six different experiments. The truth effect was measured and compared across neutral and delusion-relevant trivia information, self-referent and other-referent descriptive statements, and trait schizotypy (i.e., psychosis-proneness). Firstly, we sought to discover whether the truth effect might behave differently across different types of content, whereby delusion-relevant themes or self-relevant content might elicit greater increased perceptions of truth with repetition. Secondly, we aimed to determine whether the truth effect might behave differently across individuals, whereby those who were more delusion-prone or higher in positive schizotypy might demonstrate a larger truth effect than those lower in delusion-proneness. The key themes that emerged across the six experiments were:

1. The truth effect may not be as robust against experiment al procedures as previously suggested
2. Delusion relevant or self-relevant content type does not independently increase the magnitude of the truth effect.

3. Individual differences in trait schizotypy – specifically, positive schizotypy – predicts the size of the truth effect.
4. There appears to be an interaction between trait schizotypy and content type, whereby higher levels of positive schizotypy (as measured by the O-LIFE) predict a greater truth effect for delusion-relevant content
5. The preliminary study findings of Moritz et al. (2012) were ultimately not replicated in the present research. That is, neither content type nor CAPE positive symptom scores predicted the truth effect across our experiments.

Over this chapter, we will discuss these findings in more detail, and suggest the implications and future directions posed by the present research.

The truth effect

A surprising and unanticipated finding in the present thesis was the lack of robustness across differences in experimental procedures. Despite the extreme care and careful consideration given to developing the experimental procedure, no truth effect was observed in our first study. Previous research has stated that the truth effect is reliable both within items (d = .39; Dechêne et al., 2010) and between items (d = .49; Dechêne et al., 2010); and is robust against variations in delay (e.g., Unkelbach, 2007). However, the present research did not find a truth effect with a small delay (see Experiments 1 and 5), and only by changing the type of judgement required in response to the first round of item presentation (e.g., 'interestingness' instead of 'truth') di we find a significant effect (see Experiment 2). Furthermore, the findings of Experiment 5 – though admittedly procedurally similar to Experiment 1 – failed to show a truth effect for self-relevant statements, either providing initial evidence that the truth effect does not occur for self-relevant statements, or providing additional support for the importance of delay in facilitating illusory truth. Ultimately, we must begin this final discussion by

acknowledging that the truth effect might not be quite as robust and replicable as previously thought.

Content type and the truth effect

Several experiments over the course of this investigation explored the role of content on the truth effect. Specifically, the effect of repetition on truth was compared across delusion-relevant and neutral content (Chapters 2 and 3) and self-referent and other-referent content (Chapter 5). Ultimately, it was determined that the statements consisting of delusion-relevant content or form (self-referential) were no more subject to the truth effect than control statements. The self-referential format exerted no impact on the truth effect relative to other-referenced statements, and delusion relevance actually slightly reduced the magnitude of the observed truth effect relative to neutral trivia statements. One interpretation of this pattern is that the metacognitive sensation of processing fluency is not impacted by statement type. Processing fluency is the proposed mechanism by which statement repetition is thought to increase truth ratings and thereby elicit a truth effect (Reber & Unkelbach, 2010; Unkelbach, 2007). Therefore, the absence of an effect of statement content implies that processing fluency occurs independently of the content being evaluated, when all other variables (e.g., familiarity, individual differences, and memory effects) have been controlled for.

Delusion-relevant vs neutral trivia

While no meaningful differences were observed between delusion-relevant and neutral trivia claims and the truth effect, the present work elicited several important insights that should guide future research. First, that memory of initial ratings and a desire to be consistent can easily result in a null effect. The present experiments assumed the truth effect was robust against variations in delay between repetitions (as reported by Dechêne et al., 2010; Unkelbach,

2007). However, the only significant truth effects observed in the present experimental series occurred when the question was altered from the first repetition (interestingness) to the second (truth). The null effect for repeated truth questions was observed across content types.

Second, it appeared that familiarity played a large role in judgements of truth, such that familiarity primarily drove higher truth ratings for *novel* material. We also noticed that a proportion of delusion-relevant items were perceived as being more familiar than neutral items in Chapter 3, thus increasing baseline ratings and reducing the measured effect of repetition on truth for these statements. This likely accounted for the trend whereby neutral statements elicited a slightly larger truth effect across some of our experiments, as differences in statement type primarily occurred at baseline, rather than following repetition. However, it was also true that in our most tightly controlled experiment (Experiment 5) in which novel ratings for delusion-relevant and control materials were equated, a significantly smaller truth effect was observed for delusion-relevant than neutral content. This was not attributable to differences in novel ratings. While the precise reason for this reversed finding is not yet clear, it is evident that delusion-relevant content did not *increase* truth effect magnitude.

Self-referent vs other-referent statements

Similarly, in Chapter 5, no difference in the truth effect was observed between the two statement conditions: self-referent and other-referent material. However, unlike the comparisons between trivia statements in previous chapters, there was no difference in perceptions of truth between novel and repeated referential statements at all. Ultimately, no truth effect was observed for referential statements, suggesting that other factors play a greater role in evaluations of truth for this type of information than repetition-induced fluency. The most likely candidate here is explicit memory effects whereby participants sought to be consistent across repeated judgments for the same item. Chapter 2 revealed this drive for

consistency to play a large role in our experiments addressing delusion-relevant content, so it appears likely this factor also influenced the highly similar experiment concerning self-referential statements. Further, because self-referential statements are likely to have larger explicit memory representations than impersonal trivia (Rogers et al., 1977), the stimuli used in this experiment may have been particularly exposed to the impacts of explicit memory, as distinct from intuitive judgment. However, other factors may also have importantly reduced the magnitude of the observed truth effect in this task. Small imbalances that survived our experimental control processes in stimulus familiarity or valence, or non-linear impacts of individual differences, may also have reduce the magnitude of the observed effect. In any case, these questions are reserved for future research.

Trait schizotypy and the truth effect

In Chapter 4, the influence of individual differences in trait schizotypy (or psychosis proneness) on the truth effect was examined. This analysis utilised data from Experiments 1, 2 and 5, resulting in a robust sample size of 302 participants. The findings of these analyses revealed a small but reliable relationship between positive trait schizotypy and the truth effect, whereby the unusual experiences subscale in the O-LIFE predicted the size of the effect.

This finding was consistent with our initial predictions and provides some evidence for individual differences in susceptibility to the truth effect. To date, there have been no studies demonstrating a reliable variation in the truth effect as it relates to trait schizotypy in a healthy population, although variations in truth ratings have been explored as a function of other individual difference constructs, such as self-reported schizophrenia or the need for cognition (e.g., Moritz et al., 2012; Newman et al., 2020). However, there is more to discuss regarding the relationship between positive schizotypy and the truth effect here, as the type of content being evaluated also played a role in this finding. Furthermore, the relationship between the

truth effect and positive schizotypy emerged only on the O-LIFE measure, but no subscale in the CAPE measure of delusion-proneness was associated with truth effect magnitude. These two observations are sequentially detailed and discussed below.

Schizotypy, content type, and the truth effect

Recall that the concept linking stimulus repetition with subsequent high ratings of truth is processing fluency. Processing fluency is manipulated by variables such as repetition, which increases fluency, and is then perceived as a correlate of truth (Unkelbach, 2007). From this perspective, the present data demonstrated a greater experience of processing fluency in judgements of truth among people who are more delusion-prone, but only for content that is more analogous to delusional beliefs, such as persecutory or conspiracy themes. This is quite an unusual claim, and warrants closer investigation. To this end, the next two sections present a careful consideration of the specific relationships CAPE and O-LIFE held with the present experimental data – both in novel rating and mean truth effects – and how this might account for differences in the magnitude of their relationship with the truth effect.

Schizotypy, delusion-related material and novel truth

Before considering the relationship of each scale with the truth effect directly, note that the positive symptom subscales on both measures (unusual experiences in the O-LIFE and positive symptom frequency in the CAPE) predicted belief in delusion-relevant statements (Chapter 4) and self-referential statements (Chapter 5) when those statements were novel (i.e., not repeated). This finding revealed that positive schizotypy as a trait appears to influence perceptions of truth for content that is self-referential (but not other-referent) and delusion-relevant (but not neutral), and that this relationship is independent of processing fluency as manipulated by repetition.

While there have been previous reports of a relationship between schizotypy and belief in conspiracy theories (Darwin et al., 2011; Denovan, 2020), and fake news (Anthony & Moulding, 2019; Escolà-Gascón et al., 2023), this is the first detailed exploration of the relationship between schizotypy and belief in information related to clinical delusions specifically. Peters and colleagues (1999) have quantified the frequency and intensity of people's beliefs in themes commonly observed in delusions, but these investigations were primarily conducted to establish baseline frequencies of these beliefs for the purposes of scale formation or normalization in cognitive therapy. Similarly, scales measuring constructs similar to schizotypy have been used to predict people's belief in parapsychological phenomena, such as extra sensory perception (see the Australian Sheep-Goat Scale; Drinkwater at al., 2018). Moreover, when developing modern schizotypy scales, like the O-LIFE, earlier investigations focused closely on personality constructs associated with magical thinking (e.g., Eckblad & Chapman, 1983), which could be thought to underlie some of the proposed relationships in our delusion-relevant statement types.

Yet all of these findings are supportive of the conclusions, rather than demonstrative. The relationship between belief in these delusion-relevant or self-evaluative statements and the schizotypy construct has not been directly investigated, to our knowledge. Thus, the observed relationships between schizotypy and endorsement of items related to supernatural belief, contamination beliefs, and self-referential statements likely constitutes a new contribution to the literature in and of itself.

It is interesting to note that positive schizotypy predicted truth for delusion-relevant material exclusively, and that there was not a general tendency to rate all statements as more true overall. That is, neutral and other-referent content did not share a relationship with trait schizotypy. Equally, delusion-relevant and self-referential statements were not associated with

all subscales of the O-LIFE and CAPE measures, but instead associated primarily with their positive dimensions. Therefore, there is something about this type of information that elicits a greater perception of truth specifically. While we cannot determine the exact conditions that cause this preference for belief in delusion-relevant content with the present data, we have noted some possible contributors in previous chapters – including familiarity effects, and a preference for intuitive reasoning when evaluating delusion-like thematic content. Other accounts which have attempted to explain the relationship between positive schizotypy and endorsement of conspiracy beliefs have identified cognitive factors which play mediating or moderating roles, including intuitive thinking style and cognitive insight (Barron et al., 2018). However, a more detailed analysis of this finding is outside the scope of this thesis. The present experiments were designed to minimize and control for differences in novel ratings rather than to probe them directly. The present data merely point to likely relationships between these personality constructs and belief in themes conceptually similar to common delusional themes, and these relationships appear to exist outside of our experimental manipulations. Future research geared towards understanding the kinds of beliefs that are associated with schizotypy or delusion proneness would do well to include the current statement categories.

We turn now to the question of the association of the effect of repetition on truth with individual difference factors. This analysis focused on the delusion relevant and neutral content examined in Chapters 2-4. The self-referential statements tested in Chapter 5 are excluded because no mean truth effect was observed, so it is hard to draw any conclusions about changes in the magnitude of that effect between individuals.

In the primary analysis of the association between the OLIFE and CAPE scores and the magnitude of the truth effect, a complex hierarchical regression analysis was used. This technique allowed us to control for novel truth ratings before entering our predictor variables

(i.e., schizotypy subscale scores) for analysis. Given that novel truth ratings for delusion-relevant items shared a significant relationship with positive trait schizotypy, one might predict that positive schizotypy would either not predict the truth effect at all, or rather, that it might predict a *negative* relationship with the magnitude of the truth effect for these items. This is because highly rated novel statements allow less room for increases in endorsement following repetition. Fortunately, this was not the case. In fact, the significant relationship between the truth effect for delusion-relevant content and positive schizotypy as measured in the O-LIFE persisted *despite* the relationship between positive schizotypy and novel truth. As this finding survived rigorous statistical analysis, including the control of novel ratings, we can be confident that it is reliable and valid. Furthermore, while the effect size of this relationship was small, this can be explained by the increased baseline truth ratings for delusion-relevant items among this population. In sum, positive schizotypy – or delusion-proneness – predicts the truth effect *over and above* the tendency to endorse delusion-relevant content as being true.

This finding is indicative of an increased experience of – or reliance upon – processing fluency in judgements of truth among individuals higher in positive schizotypy. It suggests an influence of intuitive reasoning in the development of conviction for delusion-like ideas, which could quite possibly be applied to the development of clinical delusions in psychotic disorders. While we did not recruit or test a clinical sample across our investigation, there is strong evidence and consensus for a continuum model of psychosis (van Os et al., 2009), which recognises the relationship between trait schizotypy, psychosis-proneness, psychosis-risk, and schizophrenia spectrum disorders as existing along a continuum. Importantly, individuals high in schizotypy are at a higher risk of developing psychosis, and therefore clinical delusions (Grant et al. 2018). Positive schizotypy is also indicative of delusion-proneness (van Os et al., 2009). Therefore, the present findings suggest a potential role of intuitive or metacognitive

factors – specifically, processing fluency experiences – in the development of conviction among delusion-prone individuals. Specifically, the higher a person is in delusion-proneness, the more their endorsement of certain beliefs are facilitated by processing fluency.

Yet this relationship was exclusively observed for the O-LIFE and not the CAPE. This is surprising given their conceptual overlap, and indeed their significant correlation in our study sample ($r > 0.6$). The next section considers differences between the two scales that might account for the present discrepancy.

A comparison of the O-LIFE and CAPE

Given the tight correlation between the positive dimensions of the O-LIFE and CAPE, the most obvious possibility was that a collinearity effect occurred, whereby one variable simply consumed the variation explained by the other. However, this was ruled out when collinearity metrics revealed nothing of concern in Chapter 4. We turn now to a consideration of the scales themselves: their composition, their sensitivity, and their purpose. Perhaps these differences might help explain why one predicted the truth effect where the other did not.

Scale differences in measuring schizotypy

Both positive symptom frequency and unusual experiences were designed with similar purposes: to capture anomalous perceptual and cognitive experiences. While the CAPE is newer, both measures have been highly utilised in psychological research, and both have strong evidence to support their validity and reliability across time (Mason & Claridge, 2006; Stefanis et al., 2002). However, there are potentially subtle but important differences in the aspects of positive schizotypy that each scale prioritises.

First, there are clear differences in how these two measures were developed and composed. Indeed, these differences are representative of a schism in the academic literature

concerning the development, interpretation, and use of the schizotypy construct (see Grant et al., 2018, for a discussion). For instance, measures such as the O-LIFE acknowledge the fully dimensional model of psychosis, and are composed to capture the full spectrum of schizotypal traits, including nonclinical, healthy presentations. Developers of these measures recognise schizotypy, schizotypal personality disorder, and psychotic disorders such as schizophrenia as falling under the same conceptual umbrella in terms of a cluster of tendencies, traits, and symptoms (Lezenweger, 2015). Conversely, the CAPE is a measure primarily concerned with capturing subclinical psychopathology, and was developed as an evolution from existing measures of schizotypy using clinical symptoms of psychosis as a template (Stefanis et al., 2002). The CAPE is often used as a measure of psychosis and is used in research to detect probable diagnoses of psychotic disorders (e.g., Jaya et al., 2021).

Regardless of whether each scale is intended for clinical populations or for measures of general trait schizotypy, there are clearly large overlaps in the phenotypes each measure claims to capture. This overlap was again observed here, with the strong and significant correlation between scales of $r = 0.64$. The more pressing issue here is understanding the extent to which clinically focused and personality focused scales are similar, and whether there might be important qualitative differences between the scales which could explain the present differences in predictive power over truth effect magnitude.

A recent study may help answer the first part of this inquiry. Pfarr et al. (2023) asked whether there might be qualitative differences between trait schizotypy scales and prodromal psychosis screening measures. The study used Principal Component Analysis (PCA) and Confirmatory Factor Analysis (CFA) to evaluate and test the psychometric properties of a series of schizotypy scales and screening instruments for prodromal schizophrenia. Their findings provided further support for the typical three-factor structure of schizotypy (i.e., the

presence of positive, negative, and disorganised features), and found no significant qualitative or quantitative differences between prodromal psychosis and trait schizotypy measures. However, the authors agreed that, while qualitatively and quantitatively similar, different measures capture the various facets of schizotypy differently: despite significant overlap, each scale is not identical in its scope. Indeed, the present findings support this notion, as scores on the CAPE and O-LIFE were largely correlated but fell far short of being identical.

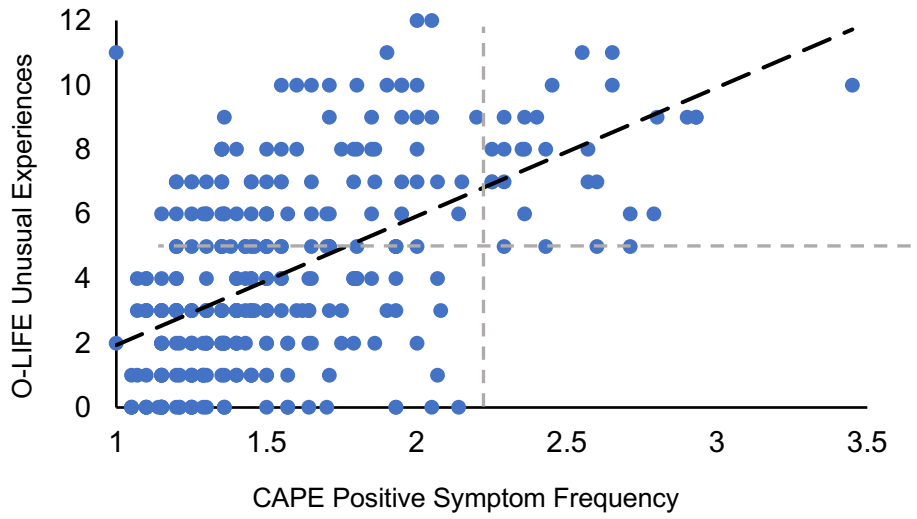
Upon closer examination of individual items in the O-LIFE and CAPE, one can notice the generality of the former and clinical relevance of the latter (see Appendix C for the full list of items in each measure). For example, the item, "Do you ever hear voices when you are alone?" in the CAPE captures an experience commonly observed in schizophrenia (i.e., auditory hallucinations), while the item "Do you ever feel as if a double has taken the place of a family member, friend or acquaintance?" is directly referring to the experience of Capgras delusion, a relatively infrequent delusion seen in individuals with a psychiatric illness (Ellis & Lewis, 2001). In contrast, the item, "Do you feel that your accidents are caused by mysterious forces?" in the O-LIFE is more relevant for capturing magical ideation, a personality feature commonly observed in the general population. While the abovementioned CAPE items are indicative of subclinical or clinical symptoms, the O-LIFE items clearly capture less clinically relevant aspects of trait schizotypy. Aside from the constructs captured by the scale items, the immediately apparent clinical purpose of the CAPE as compared with the relatively general themes of the O-LIFE may change how participants engage in the scale. For example, if one was concerned about giving a favourable (or unfavourable) clinical impression, then this bias might impact more on the apparently clinically relevant CAPE than on the O-LIFE.

However, our hypothesis is that the primary difference in the measures with respect to detecting associations with the truth effect is not subtle qualitative differences or the mental

stance of the participant, but instead lies in the relative sensitivity of the two scales to different regions of the psychosis-proneness continuum. We argue that the scope of the O-LIFE captures a lower spectrum of trait schizotypy, as it was designed to capture normal variation in personality, while the CAPE preferentially captures those same traits at a higher level that is closer to clinical disruption. That is, the CAPE may be a more conservative measure. If so, then we might expect high CAPE scores to strongly (and perhaps exclusively) predict high O-LIFE scores, but not the reverse. This hypothesis was tested by exploring the distribution of the relationship between scores on O-LIFE and CAPE, summarized in the scatter plot in Figure 17. In summary, we found substantial support for our prediction. Participants who scored highly in positive symptom frequency ($\geq$ 2.15) also reported higher scores in unusual experiences ($\geq$ 5), while participants with an unusual experiences score of 5 or higher varied greatly in their scores in positive symptom frequency. This is most directly evident in the contrast between the upper right and lower left quadrants. The upper left is highly populated (high scores on O-LIFE but low scores on CAPE) whereas the lower right quadrant (high scores on CAPE but low scores on O-LIFE) was almost empty. This suggests that the CAPE, as discussed, might capture positive schizotypy at higher levels, closer to clinically relevant levels of psychosis risk or psychosis-proneness.

Figure 17

Scatterplot (with trendline) showing the relationship between scores in positive symptom frequency as measured in the CAPE (x-axis) and unusual experiences in the O-LIFE (y axis).



Note: The grey dotted lines indicate a level at which positive symptom frequency becomes high enough (≥ 2.15) that unusual experiences is also high (≥ 5).

Scale differences in predicting the truth effect

Subtle differences in how the CAPE and O-LIFE capture schizotypy might therefore be responsible for the differences in predicting the magnitude of the truth effect in the present research. However, exactly how or why these differences might have affected our results is unclear. One explanation is that the truth effect behaves differently across schizotypy, with positive schizotypy predicting the size of the effect up to a certain point, while much higher levels of the trait (e.g., closer to clinical presentations) no longer being predictive. Another explanation lies in the possibility that participants may have underreported in the positive symptom frequency scale of the CAPE measure for fear of revealing or acknowledging perceived clinical or subclinical symptoms, impacting the validity of our results. Both explanations allow us to remain loyal to the dimensional model of psychosis, and are further outlined below.

Our first explanation follows the idea that both the CAPE and O-LIFE are qualitatively alike but differ in the level of schizotypy captured. Perhaps, in general, positive schizotypy predicts susceptibility to the truth effect (as shown by the relationship between unusual experiences and the truth effect); however, this relationship might only exist up to a certain level of schizotypy, with much higher levels of the trait no longer being predictive (as shown by the nonsignificant relationship between positive symptom frequency and the truth effect). Essentially, our theory is that a disruption or aberration in the metacognitive experience of processing fluency is related to psychosis-proneness, and this leads to larger increases in perceptions of truth following repetition. Perhaps individuals with very high levels of metacognitive aberrations – say those nearing a clinical diagnosis or those at high risk of psychosis – are not as responsive to external cues of truth when making judgements, and therefore do not experience external manipulations of processing fluency in the same way. That is, at higher levels of psychosis proneness, there may be more *internal* metacognitive aberrations occurring, which in turn exert their own influence on perceptions and judgements. An effect like this might be illustrated by a bell curve distribution, with those scoring lowest and highest in positive schizotypy being less susceptible to external manipulations of processing fluency, and those in the middle the most influenced. Of course, as this explanation does not yet have empirical evidence beyond the present findings, this is merely a hypothesis. Further research is needed to elucidate whether schizotypy influences judgements and decision-making differently across the spectrum.

Our second explanation is that some people with high experiences of psychotic-like symptoms simply underreported in the CAPE. It is true that scores in the CAPE have been demonstrated as significantly associated with objective measures of psychosis, and can be used as valid indicators of persons at ultra-high risk of a psychotic disorder (Mossaheb et al., 2012).

However, there are some items that have been shown to be dependent on an individuals' level of insight: specifically, items relating to persecutory delusions in the positive dimension of the measure (Liraud et al., 2004). Poor insight regarding the presence, severity and disruptive impact of clinical symptoms is commonly reported among people with psychotic disorders (Amador et al., 1998; Mintz et al., 2003), and this was reflected in the abovementioned study. However, the evidence for this phenomenon within subclinical or at-risk populations is less clear. On the one hand, some research has found poorer levels of insight in subclinical or at-risk populations (e.g., Kimhy et al., 2014; Xu et al., 2021), while others have reported no such effect (e.g., Dondé et al., 2021). Essentially, it appears that the question of insight is a complex one, likely varying greatly within both clinical and nonclinical populations, and across specific operationalisations of the "insight" construct.

In any case, given the high average positive CAPE scores in the present population sample, it is possible that individuals with clinically significant levels of psychosis symptomatology were present in these studies. Therefore, it remains possible that some of these individuals did have poorer levels of cognitive insight, and consequently may have underreported in response to questions relating to persecutory delusional beliefs. This would be particularly challenging from a quantification perspective, as it implies that people with a low level of the construct report low scores, those with a middle level report high scores, and those with the highest level of the construct (and thus disrupted insight) actually report low scores. Any such nonlinearity would significantly impact the scale's utility in analysis formats that assume linearity, like correlation and regression. Ultimately, however, this explanation remains highly speculative and is perhaps contradicted by the effective use of the scale to predict clinically-relevant variables in other experiments (Mossaheb et al., 2012), albeit typically with smaller sample sizes.

Comparisons with Moritz et al. (2012)

The present work developed and extended a preliminary study by Moritz et al (2012) focused on the same core question. Moritz et al. (2012) examined whether people with self-identified schizophrenia might be more prone to rely on processing fluency in judgements of truth, by conducting a truth effect experiment. More specifically, they wanted to test whether positive symptoms predicted the truth effect among people with psychosis. They also wanted to explore whether certain types of trivia statements – namely, those with emotive thematic content similar to delusions – would be more susceptible to the effects of repetition on perceived truth. Their findings were that the truth effect for delusion-relevant items was larger among those with a probable diagnosis of schizophrenia, and that scores in positive symptom frequency (as measured by the CAPE questionnaire; van Os et al., 1999) predicted the size of the truth effect for these items. It is therefore surprising that we found no overall content modulation of the truth effect, nor any relationship between the CAPE measure and truth effect magnitude.

In understanding the differences between our experiments and that of Moritz et al. (2012), it is pertinent to highlight that the present experiments were supported by the self-identified limitations in Moritz et al.'s (2012) published work. Indeed, the present experiments were designed, in part, to address these limitations (see detailed discussion in Chapters 1 and 2). The present multi-year project also had the luxury of being able to try more than one stimulus set, and thereby develop and refine the stimuli to more cleanly capture variation in delusion-relevance, as distinct to emotive content, statement length or familiarity. Our stimulus list was also longer (96 items) and more structured, with approximately equal representation over a set of delusion-like themes and content (and equivalently so for the neutral control items). With the benefit of these improvements and refinements, no effect of delusion-

relevance on the overall truth effect were observed. Where possible, analyses were repeated using only those items shared with Moritz et al.'s (2012) stimulus set. Even in these analyses we were unable to replicate their core findings, but it is equally important to note that regionally specific items were removed. So, these "matched" analyses took place with a smaller and possibly non-representative item set, and so did not constitute a direct replication.

Another important difference between the present research and that of Moritz et al. (2012) lies in the inferential analysis used. In Chapter 4, the role of individual differences in schizotypy (or psychosis-proneness) was investigated using hierarchical regression. This controlled for any joint effect of subscale of the O-LIFE or CAPE, and allowed us to probe the individual contribution of each subscale. Further, it importantly controlled for any effect of schizotypy on the ratings for novel items used in the calculation of the truth effect. This proved to be important because schizotypy predicted ratings for novel items in the delusion relevant and self-referential categories (but not the control categories). Perhaps most importantly, using a regression-based approach obviates the need to dichotomize the predictor variable and thereby captures all variance within that variable, rather than artificially inflating the differences between some scores and down-weighting the differences between others (a natural consequence of threshold-based techniques).

Using analyses that allow the preservation of continuous variables has been argued to be more informative (via retaining individual-level variability), have greater power, higher accuracy, and more efficient effect size estimates (Rucker et al., 2015). For these reasons, analyses such as linear regression are considered the preferred approach to analysing continuous variables and are highly recommended wherever possible (Comrey, 1978; MacCallum et al., 2002). As discussed, we did find a relationship between delusion-proneness and the truth effect; however, there was no relationship between scores in the positive

dimension of the CAPE and truth ratings for repeated items. Again, these findings do not provide additional support for those reported by Moritz et al. (2012).

Final analysis: Moritz et al. (2012)

While the thorough investigations and careful data analysis point to no relationship between scores in the positive dimension of the CAPE and the delusion-relevant truth effect, it is recognised that we did not conduct the identical analysis as that reported in Moritz et al.'s (2012) paper. Therefore, one final analysis to compare our results against the original findings was conducted. In the preliminary investigation by Moritz et al., scores in the positive dimension of the CAPE were split into two groups: high and low. Only those with a probable diagnosis of schizophrenia (n = 36) were included in this splitting of CAPE scores, while healthy participants were in their own category (irrespective of positive symptom frequency score). The authors then carried out a mixed ANOVA containing statement type, repetition status, truth status, and participant group. There was no indication of how these scores were split, nor how many participants remained in each category once this had occurred.

Specifically, a 2 (repetition: repeated, not repeated) x 2 (statement type: neutral, delusion-relevant) x 2 (positive trait schizotypy: low, high) mixed, repeated measures ANOVA was carried out to investigate whether our findings resembled those of Moritz et al. As truth status was not found to have a significant effect on truth and was no further reported on, we have excluded this variable from our analysis. Additionally, as we did not recruit a clinical population or ask participants about their psychiatric history, we categorised all participant scores as being high or low. Finally, as no indication as to how these scores were split in the original paper but group sizes were approximately equal, we inferred a median split was used and therefore used that here to separate the data.

Results. Data from 302 participants across Experiments 1, 2 and 5 were used in this analysis. The median of CAPE positive dimension scores was 1.35. A total of 25 participants scored 1.35 in this subscale, and so these data were excluded from our analysis. This left a remaining 135 in the high positive symptom group, and 138 in the low positive symptom group (n = 277 in total).

The results revealed a significant main effect of repetition, showing a medium effect, $F(1,275) = 54.52$, $p < .001$, $\eta^2 = .05$. This means that a truth effect was observed, as we would expect. The main effect of statement type was not significant, showing no difference in overall truth ratings between delusion-relevant and neutral items, $F(1,275) = 1.97$, $p = .16$, $\eta^2 = .001$. There was no significant interaction observed between repetition and positive schizotypy, indicating that the overall truth effect did not differ between those high and low in positive schizotypy, $F(1,275) = 3.30$, $p = .07$, $\eta^2 = .003$. However, the interaction between statement type and repetition was significant, showing a small effect, $F(1,275) = 6.52$, $p = .01$, $\eta^2 = .002$. Furthermore, the interaction between statement type and positive schizotypy group was also significant, indicating that truth ratings for delusion-relevant and neutral items differed depending on positive schizotypy group, $F(1,275) = 6.31$, $p = .01$, $\eta^2 = .003$. Finally, and most importantly, the three-way interaction between repetition, statement type, and positive schizotypy was not significant, $F(1,275) = .40$, $p = .53$, $\eta^2 < .001$.

Post-hoc analyses were conducted to understand the pattern and direction of results that are relevant to the present analysis. Firstly, a t-test with Bonferroni corrections confirmed that a truth effect occurred, and there was in fact a significant difference between novel and repeated item truth ratings, $t(275) = 7.38$, $p < .001$, $d = .46$. However, post-hoc analyses of the interaction between statement type and repetition showed no meaningful effects: no significant differences were observed between delusion-relevant and neutral truth ratings when novel,

$t(275) = 2.63$, $p = .05$, $d = .15$, nor when repeated, $t(275) = .35$, $p = 1.00$, $d = .02$. Secondly, post-hoc t-tests revealed the pattern of results behind the significant interaction between positive schizotypy and statement type observed. These showed that there was no significant difference in truth ratings between high and low positive schizotypy for neutral items, $t(275) = 1.9$, $p = .33$, $d = .19$; however, there was a significant difference in truth ratings between high and low schizotypy for delusion-relevant items, $t(275) = 4.40$, $p < .001$, $d = .43$. Specifically, those in the high positive symptom group rated delusion-relevant items as significantly more true ($M = 3.92$, $SD = .49$) than those in the low positive symptom group ($M = 3.66$, $SD = .49$). However, this was not reflected in a larger truth effect per se, and could instead reflect differences between high and low schizotypy individuals in their novel ratings for delusion relevant and neutral items, which was observed repeatedly in the present experimental series.

Even under these more favourable circumstances, we were at best only partially able to replicate Moritz et al.'s (2012) findings. Delusion-relevant content does not elicit a larger truth effect than neutral content, unlike that observed by Moritz et al. (2012). Further, positive symptom frequency was not a significant predictor of the truth effect, in opposition to Moritz et al.'s (2012) findings, which provides further confirmation for the conclusions reached in Chapter 4.

The findings of Moritz et al (2012) were ultimately not replicated in the present thesis. That is, after careful composition trivia statements, recruiting a much larger sample, and addressing the limitations outlined by Moritz et al. in their preliminary study, the truth effect for delusion-relevant items was not larger than that of neutral items, and the CAPE did not predict truth effect magnitude.

Implications

The overall contribution of this thesis is the provision of novel evidence for a more pronounced truth effect among delusion-prone individuals. Specifically, it proposes that processing fluency, as manipulated by repetition, is related to higher endorsements of truth – and thus greater conviction – for content related to delusional beliefs among people high in trait schizotypy (i.e., delusion-proneness). Furthermore, as this reliance on processing fluency reflects the use of intuitive reasoning processes, the findings suggest a relationship between the use of intuitive reasoning and belief formation.

Processing fluency, conviction, and intuition

The idea that intuitive reasoning processes might play a role in the development of delusional beliefs has been posed before (e.g., Freeman et al., 2012; Freeman et al., 2014; Ward & Garety, 2019), and has provided support for the preliminary findings reported by Moritz et al. (2012) to this point. For example, the use of intuitive reasoning strategies has been demonstrated by a pronounced reliance on certain heuristics in individuals with schizophrenia, including the availability and representativeness heuristics (Balzan et al., 2012). The cognitive mechanism thought to underlie this use of intuitive processes has been termed *hypersalience of evidence-hypothesis matches* (Speechley et al., 2010), which posits that people with psychosis may be hypersensitive to matches between preexisting beliefs and available evidence. This idea emerges from the theory of aberrant salience proposed by Kapur (2003), which suggests that disruptions in dopaminergic activity in the mesolimbic pathway (known as the *dopamine hypothesis of schizophrenia*; Meltzer & Stahl, 1976) results in an aberrant assignment of salience to internal and/or external objects, and that this underlies the development of delusions in schizophrenia. The hypersalience of evidence-hypothesis matches is thought to reflect this aberrant salience, resulting in an overreliance on weak evidence or

connections, and an early cessation of data gathering and analytical reasoning (Balzan et al., 2012).

In the context of the present findings, an overreliance on processing fluency in judgements of truth might also reflect this hypersalience of evidence-hypothesis matches, whereby the experience of relative fluency is more salient for individuals who are more delusion-prone or higher in positive schizotypy. Put simply, it might be that the match between preexisting beliefs or experiences (as demonstrated by the initial encounter and endorsement of a claim) and available evidence (the repetition of that claim as it is presented later in the experiment) is more salient for some individuals, with this hypersalience becoming more prominent at higher levels of positive schizotypy. Relative processing fluency (i.e., the discrepancy between fluent and disfluent information, as reflected by repeated and novel information) has been argued to result in fluent information being experienced as more salient (Wänke & Hansen, 2015). Potentially, an increased experience of fluency might be somewhat analogous to, or provocative of, hypersalience for certain content in certain individuals. Given that people higher in schizotypy had a pre-existing tendency to endorse delusion-relevant items as true, the subsequent repetition of these items and the relative processing fluency experienced may have rendered these items particularly salient (or 'hypersalient'). This could explain why the bias among delusion-prone individuals for endorsing novel, delusion-relevant claims also resulted in a greater truth effect for these items.

Of course, this is just one example of how the present findings might relate to current understandings of the development and maintenance of delusional belief; we do not have empirical evidence here to link the enhanced truth effect to any single neurocognitive theory of delusion formation. What we can say is that the relationship between processing fluency,

content, and delusion-proneness observed represents another intuitive mechanism by which certainty or conviction for beliefs is developed, and prompts further investigation.

The role of the truth effect in clinical delusions

As was outlined in Chapter 1, the role of the truth effect in the development of delusions could be a simple effect of repetition. Specifically, an unusual idea formed to explain an anomalous experience, repeated over time, may increase confidence in that idea until it is held with inappropriate conviction. For example, a person looks at you on the street and appears very familiar to you, but you do not know them. For some people (e.g., delusion-prone), the attempt to explain such an occurrence might come in the form of an unlikely idea, such as, "They are following me". If these unusual experiences continue over time, the idea might be repeated, increasing levels of conviction until the thought becomes a delusion. This account explains the logical role of processing fluency, and even a role for repetition specifically, in the development of strong beliefs and/or delusions.

Of course, this theory does not explain the initial experience of fluency; nor can we definitively state that an experience of fluency is what initiates a delusion. We also cannot determine whether people who are more susceptible to the truth effect simply rely more heavily on fluency in judgements, or whether they experience an objectively higher level of fluency. Nevertheless, the present findings suggest that fluency may play a role in the development of delusions in the sense that delusion-prone individuals are more likely to endorse fluent, delusion-like ideas. This might mean that, for someone more susceptible to delusions, an initial unusual thought such as "They must be following me" is likely to reach a higher level of confidence upon repetition than someone who has very low delusion-proneness.

However, there are other possible links between the truth effect and the development of delusions that do not require repetition. While the present investigation used repetition to

elicit an experience of fluency, it could be that some other factor drives unexpected or salient fluency for persons who are prone to delusions. This could be a result of internal, brain-based anomalous occurrences, such as unusual activity in neurotransmitters (e.g., dopamine; see Meltzer & Stahl, 1976), or any number of external factors manipulating processing fluency in the environment (including repetition, colour contrast, rhyming, or perhaps other, still unknown influencers of experiential fluency). In either case, this fluency feeling, paired with the accompanying idea (e.g., "They are following me") is interpreted as an indicator of truth.

Limitations of the fluency account

Whether it is repetition that causes this increasing fluency over time, or whether it is some other factor, we can suggest that it is worth pursuing experiential fluency as a contributing factor in the development of delusional beliefs; however, we cannot claim that fluency plays a causal role here. Indeed, we must not ignore the limitations of the present research or overstate the strength of its findings. Ultimately, the association between positive schizotypy and the truth effect was relatively small, indicating that processing fluency – though a potential contributing factor – is not sufficient to explain the development of unusual or erroneous beliefs. We would expect a clearer, tighter relationship to emerge if processing fluency played a causal role. Furthermore, as we did not recruit a clinical population in any of our studies, we cannot make direct claims about the role of the truth effect in clinical delusions. It is vital, therefore, that future research in this area replicate these experiments with the appropriate populations. This might include comparing the truth effect across people with schizophrenia without delusions, individuals with delusions, and a control group. It may be that these analyses show no relationship between delusions and the truth effect; or it may reveal a much stronger relationship between processing fluency and belief in a clinical population. Finally, it is possible that experiential fluency is not involved in the development of belief at all, but is

related to another, third variable not already controlled for across our experiments (e.g., mood or trait openness). The following section further considers the directions of future research in this area.

Future directions

There are several future directions that can be taken to further investigate the role of processing fluency in perceptions of truth as it relates to the development of delusions. These include replications of our truth effect experiment, with some minor changes to better understand the relationship between trait schizotypy, delusion-relevant content, processing fluency, and the truth effect. The present research question might also benefit from a neurophysiological approach, whereby Experiment 5 could be replicated with use of electroencephalogram (EEG) to record neural responses implicated in truth evaluation (e.g., Wiswede et al., 2013). Additionally, while only briefly explored here, there might be an investigation of the potential role of self-referential information in truth judgement, and the ways in which individual differences across trait schizotypy might predict truth for different types of descriptive statements.

Delusion-relevant content

Future research might further examine the interaction between positive schizotypy and the truth effect, with a focus on delusion-relevant content. This might involve a replication of Experiment 5 with the inclusion of a clinical population to make comparisons across a broader range of the psychosis spectrum. This will allow for greater specificity in determining the role of processing fluency effects in clinical delusions, in order to compare with nonclinical and delusion-prone populations. Ultimately, while our findings point to a potential role of

processing fluency in the development of delusions, it is important to test this idea with a true clinical population.

Furthermore, to better understand the role of familiarity in initial ratings of truth, the first presentation of items could involve an evaluation of its familiarity (or whether one remembers having seen the statement before), rather than its interestingness. Although we determined in Chapter 3 that familiarity was an important predictor of novel truth, familiarity and novel truth were evaluated using different participant samples. By requesting familiarity ratings for novel items in the first phase of the experiment, the direct role of familiarity effects in novel ratings can be established; further, it can be determined whether familiarity for delusion-relevant content differs depending on psychosis-proneness.

There is also reason for the adoption of a continuous scale to capture truth ratings (rather than the categorical Likert scale), as this will allow for a greater level of specificity in the shifts in perceptions of truth from novel to second encounter. There is some emerging evidence to support the use of more contemporary self-report interfaces, such as the Swipe and Slider continuous scales, due to their greater precision and straightforward use (Fryer & Nakao, 2020). Provided that the adoption of these scales includes clear labelling and an intuitive format, they might provide more precise measurement in their capacity to capture even small changes in truth judgement over time.

Finally, Experiment 5 could be readministered using an alternative manipulation of fluency. Reber and Schwarz (1999) tested the processing fluency account of the truth effect using colour contrast to manipulate ease of processing, as opposed to the classic manipulation of repetition. Their findings demonstrated the significant role of fluency in driving illusory truth, challenging the familiarity and frequency accounts posed by researchers who had only investigated the effect of repetition. Alternating the manner in which fluency is manipulated

has the added benefit of allowing investigation of any other potential cognitive mechanisms that could incidentally be confounded with manipulations of fluency or statement content. For example, if repetition or content valence incidentally led to more attention being paid to initial statements, or to negatively valenced statements, then this may well be revealed by fluency manipulations that impact on attention and salience directly (such as visual noise or figure ground contrast, used by Reber and Schwarz, 1999). The extension of the present experimental design to other manipulations of perceptual fluency will provide further information as to whether the processing fluency account is indeed useful in explaining illusory truth as it relates to schizotypy and delusion-relevant content. If so, then we can be confident in endorsing its role over and above other variables, such as memory and familiarity effects.

Electrophysiological indices of truth evaluation

Additionally, there might be an investigation of possible electrophysiological markers of the truth effect, and whether these vary across the psychosis continuum. Neuropsychological research into schizophrenia is becoming increasingly valuable in the pursuit of understanding phenomena such as delusions. In cognitive neuroscience, research has already begun to explore the neural mechanisms involved in schizophrenia, delusions, and truth evaluation in the brain (Fletcher & Frith, 2009; Kiang & Gerritsen, 2019; Moreno, Casado, & Martin-Loeches, 2016; Wiswede et al., 2013). The most commonly studied electrophysiological component that is functionally related to conceptual processing (processing the "meaning" of a stimulus), is the N400 component. The N400 has been largely studied when investigating mismatches in semantic context in word lists or sentences, and irregularities in this response has been implicated in both schizophrenia and schizotypal personality (Del Goleto, Kostova, & Blanchet, 2016; Kiang & Gerritsen, 2019; Kuperberg et al., 2018).

Wiswede et al. (2013) examined processes of truth validation by recording participants' neural activity during a reading truth evaluation task. In an exploratory analysis, the authors identified an EEG marker which appeared to specifically reflect truth vs falsity, which presented as a *late negativity* over the centroparietal scalp region between 500-1000ms, following completion of the last word in the sentence. This component was observed for false propositions in participants who had been asked to evaluate the truth of the statement. Researchers concluded that while word processing and semantic matches/mismatches appear to be automatic (reflected by the N400 present across groups), truth validation only occurred when the task demanded it and was evident in late negativity responses.

Future research might involve a replication of the truth effect experiment in Chapter 3 (Experiment 5) while recording participants' neurophysiological activity using EEG. The aim would be to look for an ERP signature – that is, a neural stimulus-response marker that reflects evaluations of truth following presentation of a statement – focusing predominantly on the late negative deflection identified by Wiswede et al. (2013). Differences in responses to repeated and non-repeated statements, as well as individual differences in positive schizotypy or psychosis-proneness, could then be evaluated.

The exploration of a neurophysiological marker for truth and conviction could yield important theoretical insights regarding individual differences in belief susceptibility, as well as to increase our understand of the experience of truth itself. For example, is truth more of a mappable, measurable occurrence in the brain, its potential directly correlated with the degree of truth reported by an individual, or is it a result of a more idiosyncratic, post-hoc interpretation of a neurological experience that is more or less consistent across people? The sensitivity of EEG in detecting the exact moment in time whereby an electrophysiological response is evoked will further allow us to make inferences regarding the preconscious versus

conscious aspects of truth evaluation, in order to inform more effective interventions for people experiencing delusions.

Conclusion

These findings do more than to just add to the existing literature on processing fluency and the effect of repetition on truth. They go a step further by implicating processing fluency in the development of belief in delusion-like ideas among people who are delusion-prone They suggest that processing fluency plays a role in the development of belief and is a possible contributor in the development and maintenance of conviction in delusions. Ultimately, they provide a strong starting point from which further research into processing fluency and conviction in clinical delusions might be explored.

References

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

Abbott, R. L. (2000). Yourdictionary.com. *Electronic Resources Review*, *4*(8), 98-98.

Alford, B. A., & Beck, A. T. (1994). Cognitive therapy of delusional beliefs. *Behaviour research and therapy*, *32*(3), 369-380.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 370-374.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3)*,* 219–235.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569-576.

Alter, A.L., & Oppenheimer, D.M. (2008). Easy on the mind, easy on the wallet: The roles of familiarity and processing fluency in valuation judgments. *Psychonomic Bulletin & Review, 15*(5), 985–990. https://doi.org/10.3758/PBR.15.5.985

Amador, X. F., & Kronengold, H. (1998). The description and meaning of insight in psychosis. *Insight and psychosis*, 15-32.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA. https://doi.org/10.1176/appi.books.9780890425596.

Ammons, R. B., & Ammons, C. H. (1962). Ammons Quick IQ Test. *Psychological Test Specialists*.

Andreasen, N. C. (1982). Negative v Positive Schizophrenia. *Archives of General Psychiatry*, *39*(7), 789. https://doi.org/10.1001/archpsyc.1982.04290070025006

Angleitner, A., & Wiggins, J. S. (Eds.). (2012). *Personality assessment via questionnaires: Current issues in theory and measurement*. Springer Science & Business Media.

Anthony, A., & Moulding, R. (2019). Breaking the news: Belief in fake news and conspiracist beliefs. *Australian Journal of Psychology*, *71*(2), 154-162.

Appelbaum, P. S., et al. (1999). Dimensional approach to delusions: comparison across types and diagnoses. *Am J Psychiatry*, 156(12), 1938-1943.

Appelbaum, P. S., Robbins, P. C., & Vesselinov, R. (2004). Persistence and stability of delusions over time. *Comprehensive Psychiatry, 45*(5), 317-324.

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*(6), 576-605.

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, *2*(2), 81-94.

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 241-252.

Bago, B., Rosenzweig, L. R., Berinsky, A. J., & Rand, D. G. (2022). Emotion may predict susceptibility to fake news but emotion regulation does not seem to help. *Cognition and Emotion,* ahead-of-print(ahead-of-print), 1–15.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329-349.

Balzan, R., Delfabbro, P., & Galletly, C. (2012). Delusion-proneness or miscomprehension? A re-examination of the jumping-to-conclusions bias. *Australian Journal of Psychology*, *64*(2), 100-107.

Balzan, R., Delfabbro, P., Galletly, C., & Woodward, T. (2012). Reasoning heuristics across the psychosis continuum: The contribution of hypersalient evidence–hypothesis matches. *Cognitive Neuropsychiatry*, *17*(5), 431-450.

Barchetti, A., Neybert, E., Mantel, S. P., & Kardes, F. R. (2022). The Half-Truth Effect and Its Implications for Sustainability. *Sustainability, 14*(11), 6943.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, *43*(3), 425–436. https://doi.org/10.1037/0022-3514.43.3.425.

Barron, D., Furnham, A., Weis, L., Morgan, K. D., Towell, T., & Swami, V. (2018). The relationship between schizotypal facets and conspiracist beliefs via cognitive processes. *Psychiatry Research,* 259, 15-20.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037//1089-2680.5.4.323

Beck, A. T. (1979). *Cognitive therapy of depression.* Guilford Press.

Beck, J. S. (2020). *Cognitive behavior therapy: Basics and beyond*. Guilford Publications.

Beck, J. S., & Beck, A. T. (2011). *Cognitive behavior therapy: Basics and beyond.* Guilford Publications.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*(4), 446.

Begg, I., & Armour, V. (1991). Repetition and the ring of truth: Biasing comments. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *23*(2), 195.

Bentall, R. P., Corcoran, R., Howard, R., Blackwood, N., & Kinderman, P. (2001). Persecutory delusions: A review and theoretical integration. *Clinical Psychology Review, 21*(8), 1143-1192.

Bentall, R. P., Kinderman, P., & Kaney, S. (1994). The self, attributional processes and abnormal beliefs: towards a model of persecutory delusions. *Behaviour Research and Therapy, 32(3), 331–341.*

Berrios, G. E., & Luque, R. (1995). Cotard's delusion or syndrome?: a conceptual history. *Comprehensive psychiatry*, *36*(3), 218-223.

Bitter, I., Fehér, L., Tényi, T., & Czobor, P. (2015). Treatment adherence and insight in schizophrenia. *Psychiatria Hungarica: A Magyar Pszichiatriai Tarsasag tudomanyos folyoirata, 30*(1), 18–26.

Blackwood, N. J., Howard, R. J., Bentall, R. P., & Murray, R. M. (2001). Cognitive neuropsychiatric models of persecutory delusions. *American Journal of Psychiatry, 158*(4), 527-539.

Blanchette, I. (2006). The effect of emotion on interpretation and logic in a conditional reasoning task. *Memory & Cognition*, *34*(5), 1112–1125.

Blanchette, I., & Richards, A. (2004). Reasoning about emotional and neutral materials: Is logic affected by emotion? *Psychological Science*, *15*(11), 745–752.

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition and Emotion*, *24*(4), 561–595.

Bohner, G., Bless, H., Schwarz, N., & Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology, 18,* 335–345.

Bourgin, J., Tebeka, S., Mallet, J., Mazer, N., Dubertret, C., & Le Strat, Y. (2020). Prevalence and correlates of psychotic-like experiences in the general population. Schizophrenia research, 215, 371-377.

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology, 71,* 499–515. https://doi.org/10.1146/annurev-psych-010419-050807

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition, 194*, 104054–104054. https://doi.org/10.1016/j.cognition.2019.104054.

Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1088.

Broyd, A., Ettinger, U., & Thoma, V. (2019). Thinking dispositions and cognitive reflection performance in schizotypy. *Judgment and Decision Making*, *14*(1), 80–90. https://doi.org/10.1017/s193029750000293x

Buchanan, R. W. (2007). Persistent negative symptoms in schizophrenia: An overview. *Schizophrenia Bulletin*, 33, 1013–1022.

Buchy, L., Woodward, T. S., & Liotti, M. (2007). A cognitive bias against disconfirmatory evidence (BADE) is associated with schizotypy. *Schizophrenia Research, 90*(1-3), 334-337.

Burton, R. A. (2009). *On being certain: Believing you are right even when you're not*. Macmillan.

Casey, P., & Kelly, B. (2019). *Fish's Clinical Psychopathology: Signs and Symptoms in Psychiatry*: Cambridge University Press.

Claridge, G. (1985). *Origins of Mental Illness: Temperament, Deviance and Disorder.* Oxford, UK: Blackwell.

Cohen, J. D., Barch, D. M., Carter, C., & Servan-Schreiber, D. (1999). Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. *Journal of abnormal psychology*, *108*(1), 120.

Coltheart, M., Langdon, R., & Mckay, R. (2011). Delusional Belief. *Annual Review of Psychology*, *62*(1), 271–298. https://doi.org/10.1146/annurev.psych.121208.131622

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology, 46*(4), 648.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, *58*(3), 306-324.

Corcoran, R. (2003). Inductive reasoning and the understanding of intention in schizophrenia. *Cognitive neuropsychiatry*, *8*(3), 223-235.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophrenia research*, *17*(1), 5-13.

Corlett, P. R., Krystal, J. H., Taylor, J. R., & Fletcher, P. C. (2009). Why do delusions persist? *Frontiers in Human Neuroscience*, *3*, 12.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbing, T. W., … Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain, 130*(9), 2387-2400.

Corlett, P. R., Simons, J. S., Pigott, J., 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*, 965.

Cowan, H. R., & Mittal, V. A. (2021). Three types of psychotic-like experiences in youth at clinical high risk for psychosis. *European Archives of Psychiatry and Clinical Neuroscience*, *271*(4), 733–744. https://doi.org/10.1007/s00406-020-01143-w

Craig, J. S., Hatton, C., Craig, F. B., & Bentall, R. P. (2004). Persecutory beliefs, attributions and theory of mind: comparison of patients with paranoid delusions, Asperger's syndrome and healthy controls. *Schizophrenia research*, *69*(1), 29-33.

Craik, F. I. (2000). Age-related changes in human memory. *Cognitive aging: A primer*, *5*, 75-92.

Crawford, J. R., & Howell, D. C. (1998). Comparing an Individual's Test Score Against Norms Derived from Small Samples. *The Clinical Neuropsychologist, 12*(4), 482–486. https://doi.org/10.1076/clin.12.4.482.7241

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281.

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

De Keersmaecker, J., Dunning, D., Pennycook, G., Rand, D. G., Sanchez, C., Unkelbach, C., & Roets, A. (2020). Investigating the robustness of the illusory truth effect across individual differences in cognitive ability, need for cognitive closure, and cognitive style. *Personality and Social Psychology Bulletin, 46*(2), 204-215.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology, 45*(5), 1117–1122.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238-257.

Del Goleto, S., Kostova, M., & Blanchet, A. (2016). Impaired context processing during irony comprehension in schizotypy: An ERPs study. *International Journal of Psychophysiology, 105*, 17-25.

Denovan, A., Dagnall, N., Drinkwater, K., Parker, A., & Neave, N. (2020). Conspiracist beliefs, intuitive thinking, and schizotypal facets: A further evaluation. *Applied Cognitive Psychology*, *34*(6), 1394–1405. https://doi.org/10.1002/acp.3716

Dickinson, D., & Coursey, R. D. (2002). Independence and overlap among neurocognitive correlates of community functioning in schizophrenia. *Schizophrenia research*, *56*(1-2), 161-170.

Dondé, C., Laprévote, V., Lavallé, L., Haesebaert, F., Fakra, E., & Brunelin, J. (2021). Cognitive insight in individuals with an at-risk mental state for psychosis: A meta-analysis. *Early Intervention in Psychiatry*, *15*(3), 449–456. https://doi.org/10.1111/eip.12993

Drinkwater, K., Denovan, A., Dagnall, N., & Parker, A. (2018). The Australian sheep-goat scale: an evaluation of factor structure and convergent validity. *Frontiers in Psychology*, *9*, 1594.

Dudley, R., Taylor, P., Wickham, S., & Hutton, P. (2016). Psychosis, delusions and the "jumping to conclusions" reasoning bias: A systematic review and meta-analysis. *Schizophrenia Bulletin, 42*(3), 652-665.

Dudley, R.E.J., John, C.H., Young, A.W. and Over, D.E. (1997), The effect of self-referent material on the reasoning of people with delusions. *British Journal of Clinical Psychology, 36,* 575-584.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121.

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, *51*(2), 215.

Ellinas, C., Allan, N., & Johansson, A. (2017). Dynamics of organizational culture: Individual beliefs vs. social conformity. *PloS one*, *12*(6), e0180193.

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. *Cerebral Cortex*, *10*(3), 308-317.

Ellis, H. D., & Lewis, M. B. (2001). Capgras delusion: a window on face recognition. *Trends in Cognitive Sciences*, *5*(4), 149-156.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390.

Escolà-Gascón, Á., Dagnall, N., Denovan, A., Drinkwater, K., & Diez-Bosch, M. (2023). Who falls for fake news? Psychological and clinical profiling evidence of fake news consumers. *Personality and Individual Differences*, *200*, 111893.

Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, *80*(3), 532-545.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993.

Fazio, L. K., & Sherry, C. L. (2020). The effect of repetition on truth judgments across development. *Psychological Science, 31*(9), 1150–1160. https://doi.org/10.1177/0956797620939534

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin & Review, 26*(5), 1705–1710. https://doi.org/10.3758/s13423-019-01651-4

Fitzsimmons, C.J., Thompson, C.A., & Sidney, P.G. (2020). Confident or familiar? The role of familiarity ratings in adults' confidence judgments when estimating fraction magnitudes. *Metacognition and Learning, 15*(2), 215–231.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci, 10*(1), 48-58.

Fonseca-Pedrero, E., Ortuño-Sierra, J., Mason, O. J., & Muñiz, J. (2015). The Oxford–Liverpool Inventory of Feelings and Experiences short version: Further validation. *Personality and Individual Differences*, *86*, 338-343.

Fox, J. (2016). Applied regression analysis and generalized linear models (3rd ed.). Los Angeles: Sage Publications.

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

Freeman, D., Evans, N., & Lister, R. (2012). Gut feelings, deliberative thought, and paranoid ideation: A study of experiential and rational reasoning. *Psychiatry Research*, *197*(1-2), 119-122.

Freeman, D., & Garety, P. (2014). Advances in understanding and treating persecutory delusions: a review. *Social Psychiatry and Psychiatric Epidemiology, 49*(8), 1179–1189. https://doi.org/10.1007/s00127-014-0928-7

Freeman, D., Garety, P. A., Kuipers, E., Fowler, D., & Bebbington, P. E. (2002). A cognitive model of persecutory delusions. *British Journal of Clinical Psychology*, *41*(4), 331-347.

Freeman, D., Lister, R., & Evans, N. (2014). The use of intuitive and analytic reasoning styles by patients with persecutory delusions. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(4), 454–458.

Friston, K. J., & Frith, C. D. (1995). Schizophrenia: a disconnection syndrome. *Clinical Neuroscience*, *3*(2), 89-97.

Frith, C. D. (1992). The cognitive neuropsychology of schizophrenia. Hillsdale, U.S., Hove, U.K.

Fryer, L. K., & Nakao, K. (2020). The Future of Survey Self-Report: An Experiment Contrasting Likert, VAS, Slide, and Swipe Touch Interfaces. *Frontline Learning Research*, *8*(3), 10-25.

Fujita, T., & Horiuchi, T. (1998). Self-reference effect on implicit and explicit memory tasks. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *69*(5), 414-420.

Garcia-Marques, T., Prada, M., & Mackie, D.M. (2016). Familiarity increases subjective positive affect even in non-affective and non-evaluative contexts. *Motivation and Emotion, 40*(4), 638–645.

Garcia-Marques, T., Silva, R. R., Reber, R., & Unkelbach, C. (2015). Hearing a statement now and believing the opposite later. *Journal of Experimental Social Psychology*, *56*, 126-129.

García-Montes, J. M., Pérez-Álvarez, M., Odriozola-González, P., Vallina-Fernández, O., & Perona-Garcelán, S. (2014). The role of magical thinking in hallucinations. Comparisons of clinical and non-clinical groups. *Nordic Journal of Psychiatry, 68*(8), 605–610. https://doi.org/10.3109/08039488.2014.902500

Garety, P. (1991). Reasoning and delusions. *British Journal of Psychiatry, 159*(14), 14-18.

Garety, P. A., & Freeman, D. (1999). Cognitive approaches to delusions: a critical review of theories and evidence. British journal of clinical psychology, 38(2), 113-154.

Garety, P. A., & Freeman, D. (2013). The past and the future of delusions research: from the inexplicable to the treatable. *Br J Psychiatry, 203*(5), 327-333.

Garety, P. A., Fowler, D., & Kuipers, E. (2000). Cognitive-behavioral therapy for medication-resistant symptoms. *Schizophrenia Bulletin, 26*(1), 73-86.

Garety, P., Freeman, D., Jolley, S., Ross, K., Waller, H., & Dunn, G. (2011). Jumping to conclusions: the psychology of delusional reasoning. *Advances in psychiatric treatment*, *17*(5), 332-339.

Garner, D. M., & Bemis, K. M. (1982). A cognitive-behavioral approach to anorexia nervosa. *Cognitive Therapy and Research*, *6*, 123-150.

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *The American journal of psychology*, 185-195.

Gilleen, J., Tesse, M., Velikonja, T., Weiser, M., Davidson, M., & Reichenberg, A. (2020). Schizotypal traits and neuropsychological performance: The role of processing speed. *Schizophrenia Research*. doi:10.1016/j.schres.2020.06.030

Grant, P., Green, M. J., & Mason, O. J. (2018). Models of schizotypy: The importance of conceptual clarity. *Schizophrenia Bulletin, 44,* 2, S556-S563.

Griffiths, O., & Mitchell, C. J. (2008). Negative priming reduces affective ratings. *Cognition and Emotion, 22*(6), 1119-1129.

Grimmer, H. J., Laukkonen, R. E., Freydenzon, A., von Hippel, W., & Tangen, J. M. (2022). Thinking style and psychosis proneness do not predict false insights. *Consciousness and Cognition*, *104*, 103384.

Haddad, P. M., & Correll, C. U. (2018). The acute efficacy of antipsychotics in schizophrenia: a review of recent meta-analyses. *Therapeutic advances in psychopharmacology*, *8*(11), 303-318.

Haddock, G., Tarrier, N., Spaulding, W., Yusupoff, L., Kinney, C., & McCarthy, E. (1998). Individual cognitive-behavior therapy in the treatment of hallucinations and delusions: a review. *Clinical Psychology Review*, *18*(7), 821-838.

Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology, 44,* 687–691.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*(4), 208–216. https://doi.org/10.1037/h0022263

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16*(1), 107-112.

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive research: principles and implications*, *6*(1), 1-12.

Hawkins, S. A., & Hoch, S. J. (1992). Low-involvement learning: Memory without evaluation. *Journal of consumer research*, *19*(2), 212-225.

Heit, E. (2000). Properties of inductive reasoning. Psychonomic Bulletin & Review, 7(4), 569-592.

Hemsley, D. R. (1993). A simple (or simplistic?) cognitive model for schizophrenia. *Behaviour Research and Therapy*, *31*(7), 633–645. https://doi.org/10.1016/0005-7967(93)90116-c

Henkel, Linda A, & Mattson, Mark E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and Cognition, 20*(4), 1705–1721.

Hilbig, B. E. (2009). Sad, thus true: negativity bias in judgments of truth. *Journal of Experimental Social Psychology, 45*(4), 983-986.

Hilbig, B. E. (2012a). Good things don't come easy (to mind): Explaining framing effects in judgments of truth. *Experimental Psychology, 59*(1), 38–46. https://doi.org/10.1027/1618-3169/a000124

Hilbig, B. E. (2012b). How framing statistical statements affects subjective veracity: Validation and application of a multinomial model for judgments of truth. *Cognition*, *125*(1), 37-48.

Hinterbuchinger, B., & Mossaheb, N. (2021). Psychotic-like experiences: a challenge in definition and assessment. *Frontiers in Psychiatry*, *12*, 582392.

Holland, R. W., Verplanken, B., & van Knippenberg, A. (2003). From repetition to conviction: Attitude accessibility as a determinant of attitude certainty. *Journal of experimental social psychology*, *39*(6), 594-601.

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. *Q J Exp Psychol A, 40*(4), 801-812.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*(3), 306-340.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118*(2), 126-135.

Jaffé, M. E., & Greifeneder, R. (2021). Negative or negated, thus true? An investigation of concept valence and semantic negation as drivers of framing effects in judgments of truth. *Social Cognition*, *39*(6), 687-716.

Jaya, E. S., van Amelsvoort, T., Bartels-Velthuis, A. A., Bruggeman, R., Cahn, W., de Haan, L., Kahn, R. S., van Os, J., Schirmbeck, F., Simons, C. J. P., & Lincoln, T. M. (2021). The Community Assessment of Psychic Experiences: Optimal cut-off scores for detecting individuals with a psychotic disorder. *International Journal of Methods in Psychiatric Research*, *30*(4), e1893. https://doi.org/10.1002/mpr.1893.

Jenkins, A. C., & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience*, *6*(3), 211-218.

John, C., & Dodgson, G. (1994). Inductive reasoning in delusional thought. *Journal of Mental Health*, *3*(1), 31-49.

Johns, L. C., & Van Os, J. (2001). The continuity of psychotic experiences in the general population. *Clinical Psychology Review, 21*(8), 1125–1141. https://doi.org/10.1016/s0272-7358(01)00103-9.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological review*, *88*(1), 67.

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 210–223.

Jolley, S., Thompson, C., Hurley, J., Medin, E., Butler, L., Bebbington, P., . . . Garety, P. (2014). Jumping to the wrong conclusions? An investigation of the mechanisms of reasoning errors in delusions. *Psychiatry Res, 219*(2), 275-282.

Jones, C., Leung, N., & Harris, G. (2007). Dysfunctional core beliefs in eating disorders: A review. *Journal of Cognitive Psychotherapy*, *21*(2), 156.

Jones, E., & Watson, J. P. (1997). Delusion, the overvalued idea and religious beliefs: a comparative analysis of their characteristics. *The British Journal of Psychiatry*, *170*, 381.

Jones, E., & Watson, J. P. (1997). Delusion, the overvalued idea and religious beliefs: a comparative analysis of their characteristics. *The British Journal of Psychiatry*, *170*, 381.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697-720.

Kahneman, D. (2003). Maps of bounded rationality. *American Economic Review, 93*(5), 1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, Lovallo, D., & Sibony, O. (2011). Before you make that big decision. In *Harvard business review* (Vol. 89, Issue 6, pp. 50–137). Harvard Business School Publishing Corporation.

Kansal, N. K., Chawla, O., & Singh, G. P. (2012). Treatment of delusional infestation with olanzapine. *Indian Journal of Psychological Medicine, 34*(3), 297–298. https://doi.org/10.4103/0253-7176.106043

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *American journal of Psychiatry*, *160*(1), 13-23.

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of Consumer Research, 14*(2), 200-213.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of memory and language*, *32*(1), 1-24.

Kendler, K. S., Gallagher, T. J., Abelson, J. M., & Kessler, R. C. (1996). Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample: the National Comorbidity Survey. Archives of general psychiatry, 53(11), 1022-1031.

Kendler, K. S., Glazer, W. M., & Morgenstern, H. (1983). Dimensions of delusional experience. *The American Journal of Psychiatry, 140*(4), 466-469.

Kersting, M., Althoff, K., & Jäger, A. O. (2008). Wilde Intelligence Test 2 (dt. Wilde-Intelligenz-Test 2). *Göttingen: Hogrefe*.

Kiang, M., & Gerritsen, C. J. (2019). The N400 event-related brain potential response: A window on deficits in predicting meaning in schizophrenia. *International Journal of Psychophysiology, 145*, 65-69.

Kiken, L. G. (2019). Mind-body therapy use and magical thinking. *Social Science & Medicine, 237*, 112340. https://doi.org/10.1016/j.socscimed.2019.05.045

Kimhy, D., Jobson-Ahmed, L., Ben-David, S., Ramadhar, L., Malaspina, D., & Corcoran, C. M. (2014). Cognitive insight in individuals at clinical high risk for psychosis. *Early Intervention in Psychiatry*, *8*(2), 130–137. https://doi.org/10.1111/eip.12023

Kiran, C., & Chaudhury, S. (2009). Understanding delusions. *Industrial Psychiatry Journal*, *18*(1), 3-18.

Klein, S. B. (2012). Self, memory, and the self-reference effect: An examination of conceptual and methodological issues. *Personality and Social Psychology Review, 16*(3), 283-300.

Knowles, E. E., Weiser, M., David, A. S., Glahn, D. C., Davidson, M., & Reichenberg, A. (2015). The puzzle of processing speed, memory, and executive function impairments in schizophrenia: Fitting the pieces together. *Biological psychiatry*, *78*(11), 786-793.

Knowles, R., McCarthy-Jones, S., & Rowse, G. (2011). Grandiose delusions: A review and theoretical integration of cognitive and affective perspectives. *Clinical Psychology Review*, *31*(4), 684-696.

Koch, A. S., & Forgas, J. P. (2012). Feeling good and feeling truth: The interactive effects of mood and processing fluency on truth judgments. *Journal of Experimental Social Psychology, 48*(2), 481-485.

Konings, M, Bak, M, Hanssen, M, van Os, J, & Krabbendam, L. (2006). Validity and reliability of the CAPE: a self-report instrument for the measurement of psychotic experiences in the general population. *Acta Psychiatrica Scandinavica*, *114*(1), 55–61. https://doi.org/10.1111/j.1600-0447.2005.00741.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609–639.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, *9*(2), 149-171.

Kruglanski, A. W. (1990). Motivations for judging and knowing: Implications for causal attribution. In E.T. Higgins, & R.M. Sorrentinio (Eds.), *The handbook of motivation*

and cognition: Foundation of social behavior (pp. 333-368). Guilford Press, New York.

Kuperberg, G. R., Delaney-Busch, N., Fanucci, K., & Blackford, T. (2018). Priming production: Neural evidence for enhanced automatic semantic activity preceding language production in schizophrenia. *Neuroimage Clin, 18*, 74-85.

Kwapil, T. R., Gross, G. M., Silvia, P. J., Raulin, M. L., & Barrantes-Vidal, N. (2018). Development and psychometric properties of the Multidimensional Schizotypy Scale: A new measure for assessing positive, negative, and disorganized schizotypy. *Schizophrenia research*, *193*, 209-217.

Kwaśniewicz, Ł., Wójcik, G. M., Kawiak, A., Schneider, P., and Wierzbicki, A. (2020). "How you say or what you say? Neural activity in message credibility evaluation," in International Conference on Computational Science (Cham: Springer), 312–326. doi: 10.1007/978-3-030-50371-0_23

Kwaśniewicz, Ł., Wójcik, G. M., Schneider, P., Kawiak, A., and Wierzbicki, A. (2021). What to believe? Impact of knowledge and message length on neural activity in message credibility evaluation. *Frontiers in Human Neuroscience, 15,* 659243–659243. https://doi.org/10.3389/fnhum.2021.659243

Langdon, R., Ward, P. B., & Coltheart, M. (2010). Reasoning anomalies associated with delusions in schizophrenia. *Schizophr Bull, 36*(2), 321-330.

Larivière, S., Lavigne, K. M., Woodward, T. S., Gerretsen, P., Graff-Guerrero, A., & Menon, M. (2017). Altered functional connectivity in brain networks underlying self-referential processing in delusions of reference in schizophrenia. *Psychiatry Research: Neuroimaging*, *263*, 32-43.

Law, S., Hawkins, S. A., & Craik, F. I. M. (1998). Repetition-Induced Belief in the Elderly: Rehabilitating Age-Related Memory Deficits. *The Journal of Consumer Research, 25*(2), 91–107. https://doi.org/10.1086/209529

Lenzenweger, M. F. (2015). Thinking clearly about schizotypy: hewing to the schizophrenia liability core, considering interesting tangents, and avoiding conceptual quicksand. *Schizophrenia Bulletin*, *41*(suppl_2), S483-S491.

Lewicka, M. (1997). Is hate wiser than love? Cognitive and emotional utilities in decision making. In R. Ranyard, W. R. Crozier, & O. Svenson (Eds.), Decision making: Cognitive models and explanations (pp. 90–108). London: Routledge.

Li, M-.H., Chen, Z., & Rao, L.-L. (2022). Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Computers in Human Behavior, 133,* 107295–107295.

Lincoln, T. M., Ziegler, M., Mehl, S., & Rief, W. (2010). The jumping to conclusions bias in delusions: specificity and changeability. *J Abnorm Psychol, 119*(1), 40-49.

Liraud, F., Droulout, T., Parrot, M., & Verdoux, H. (2004). Agreement between self-rated and clinically assessed symptoms in subjects with psychosis. *The Journal of Nervous and Mental Disease*, *192*(5), 352-356.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19.

Maddux, J. E., & Meier, L. J. (1995). Self-Efficacy and Depression. In *The Plenum Series in Social/Clinical Psychology* (pp. 143–169). The Plenum Series in Social/Clinical Psychology. https://doi.org/10.1007/978-1-4419-6868-5_5.

Madison, G., & Schiölde, G. (2017). Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music. *Frontiers in Neuroscience, 11,* 147–147. https://doi.org/10.3389/fnins.2017.00147

Maher, B. (2005). Delusional thinking and cognitive disorder. *Integr Physiol Behav Sci, 40*(3), 136-146.

Mason, O. J. (2015). The assessment of schizotypy and its clinical relevance. *Schizophrenia bulletin*, *41*(2), S374-S385.

Mason, O., & Claridge, G. (2006). The Oxford-Liverpool Inventory of Feelings and Experiences (O-LIFE): Further description and extended norms. *Schizophrenia Research*, *82*(2), 203–211. https://doi.org/10.1016/j.schres.2005.12.845.

Mason, O., Claridge, G., & Jackson, M. (1995). New scales for the assessment of schizotypy. *Personality and Individual Differences*, *18*(1), 7-13.

Mason, O., Linney, Y., & Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophrenia Research, 78*, 293-296.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological science*, *11*(5), 424-428.

McLean, B. F., Mattiske, J. K., & Balzan, R. P. (2016). Association of the jumping to conclusions and evidence integration biases with delusions in psychosis: A detailed meta-analysis. *Schizophrenia Bulletin, 43*(2), 344-354.

McLean, B. F., Mattiske, J. K., & Balzan, R. P. (2018). Towards a reliable repeated-measures beads task for assessing the jumping to conclusions bias. *Psychiatry Res, 265*, 200-207.

Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, *17*(12), 827.

Meltzer, H. Y., & Stahl, S. M. (1976). The dopamine hypothesis of schizophrenia: a review. *Schizophrenia Bulletin*, *2*(1), 19.

Mintz, A. R., Dobson, K. S., & Romney, D. M. (2003). Insight in schizophrenia: a meta-analysis. *Schizophrenia Research, 61*(1), 75-88.

Mirian, D., Heinrichs, R. W., & Vaz, S. M. (2011). Exploring logical reasoning abilities in schizophrenia patients. *Schizophrenia Research, 127*(1-3), 178-180.

Moritz, S., & Woodward, T. S. (2006). A generalized bias against disconfirmatory evidence in schizophrenia. *Psychiatry Res, 142*(2-3), 157-165.

Moritz, S., & Woodward, T. S. (2007). Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. *Current opinion in psychiatry*, *20*(6), 619-625.

Moritz, S., & Woodward, T. S. (2005). Jumping to conclusions in delusional and non-delusional schizophrenic patients. *British Journal of Clinical Psychology*, *44*, 193– 207.

Moritz, S., Göritz, A. S., Balzan, R. P., Gawęda, Ł., Kulagin, S. C., & Andreou, C. (2017). A new paradigm to measure probabilistic reasoning and a possible answer to the question why psychosis-prone individuals jump to conclusions. *Journal of Abnormal Psychology*, *126*(4), 406.

Moritz, S., Köther, U., Woodward, T. S., Veckenstedt, R., Dechêne, A., & Stahl, C. (2012). Repetition is good? An Internet trial on the illusory truth effect in schizophrenia and nonclinical participants. *Journal of Behavior Therapy and Experimental Psychiatry, 43*(4), 1058-1063.

Moritz, S., Mayer-Stassfurth, H., Endlich, L., Andreou, C., Ramdani, N., Petermann, F., & Balzan, R. P. (2015). The benefits of doubt: cognitive bias correction reduces hasty decision-making in schizophrenia. *Cognitive Therapy and Research*, *39*, 627-635.

Moritz, S., Woodward, T.S., & Balzan, R. (2016). Is metacognitive training for psychosis effective? *Expert Review of Neurotherapeutics, 16*(2), 105-107.

Mossaheb, N., Becker, J., Schaefer, M. R., Klier, C. M., Schloegelhofer, M., Papageorgiou, K., & Amminger, G. P. (2012). The Community Assessment of Psychic Experience (CAPE) questionnaire as a screening-instrument in the detection of individuals at ultra-high risk for psychosis. *Schizophrenia Research*, *141*(2-3), 210–214. https://doi.org/10.1016/j.schres.2012.08.008

Mullen, R., & Linscott, R. J. (2010). A comparison of delusions and overvalued ideas. *The Journal of Nervous and Mental Disease*, *198*(1), 35-38.

Nadarevic, L., & Erdfelder, E. (2014). Initial judgment task and delay of the final validity-rating task moderate the truth effect. *Consciousness and Cognition, 23,* 74–84. https://doi.org/10.1016/j.concog.2013.12.002

Nahari, G. (2018). The applicability of the verifiability approach to the real world. In *Detecting concealed information and deception* (pp. 329-349). Academic Press.

Nahari, G., & Nisin, Z. (2019). Digging further into the speech of liars: Future research prospects in verbal lie detection. *Frontiers in Psychiatry*, *10*, 56.

Nahon, L. S., Teige-Mocigemba, S., Reber, R., & Greifeneder, R. (2021). Truth feels easy: Knowing information is true enhances experienced processing fluency. *Cognition, 215,* 104819. https://doi.org/10.1016/j.cognition.2021.104819

Nelson, M. T., Seal, M. L., Pantelis, C., & Phillips, L. J. (2013). Evidence of a dimensional relationship between schizotypy and schizophrenia: a systematic review. *Neuroscience & Biobehavioral Reviews*, *37*(3), 317-327.

Newman, E. J., Jalbert, M. C., Schwarz, N., & Ly, D. P. (2020). Truthiness, the illusory truth effect, and the role of need for cognition. *Consciousness and Cognition, 78,* 102866.

Newman, M. C. (2018). Human Reasoning. In *The Nature and Use of Ecotoxicological Evidence* (pp. 19-67).

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12*(6), 237-241.

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004.

Parks, C. M., & Toth, J. P. (2006). Fluency, familiarity, aging, and the illusion of truth. *Aging, Neuropsychology, and Cognition*, *13*(2), 225-253.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, *24*(6), 425-432.

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185-200.

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

Petty, R. E., & Briñol, P. (2008). Persuasion: From single to multiple to metacognitive processes. *Perspectives on Psychological Science, 3,* 137–147.

Pfarr, J. K., Meller, T., Evermann, U., Sahakyan, L., Kwapil, T. R., & Nenadić, I. (2023). Trait schizotypy and the psychosis prodrome: Current standard assessment of extended psychosis spectrum phenotypes. *Schizophrenia Research*, *254*, 208-217.

Picardi, A., Fonzi, L., Pallagrosi, M., Gigantesco, A., & Biondi, M. (2018). Delusional themes across affective and non-affective psychoses. *Frontiers in Psychiatry*, *9*, 132.

Potkin, S. G., Kane, J. M., Correll, C. U., Lindenmayer, J. P., Agid, O., Marder, S. R., ... & Howes, O. D. (2020). The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. *NPJ Schizophrenia*, *6*(1), 1-10.

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science, 357*(6351), 596-600.

Pytlik, N., Soll, D., & Mehl, S. (2020). Thinking Preferences and Conspiracy Belief: Intuitive Thinking and the Jumping to Conclusions-Bias as a Basis for the Belief in Conspiracy Theories. *Frontiers in Psychiatry, 11*, 568942–568942. https://doi.org/10.3389/fpsyt.2020.568942

Raine, A. (1991). The SPQ: A Scale for the Assessment of Schizotypal Personality Based on DSM-III-R Criteria. *Schizophrenia Bulletin, 17*(4), 555-564.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgements of truth. *Consciousness and Cognition, 8*(3), 338-342.

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of philosophy and psychology, 1*(4), 563-581.

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*(1), 47–60. https://doi.org/10.1016/S1053-8100(03)00049-7.

Riesthuis, P., & Woods, J. (2024). "That's just like, your opinion, man": the illusory truth effect on opinions. *Psychological Research*, *88*(1), 284-306.

Risen, J. L. (2016). Believing what we do not believe: Acquiescence to superstitious beliefs and other powerful intuitions. *Psychological Review, 123*(2), 182–207. https://doi.org/10.1037/rev0000017

Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. Personality and Individual Differences, 50(1), 90–94. https://doi.org/10.1016/j.paid.2010.09.004

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677.

Ross, R. M., McKay, R., Coltheart, M., & Langdon, R. (2015). Jumping to Conclusions About the Beads Task? A Meta-analysis of Delusional Ideation and Data-Gathering. *Schizophrenia Bulletin, 41*(5), 1183-1191.

Rössler, W., Riecher-Rössler, A., Angst, J., Murray, R., Gamma, A., Eich, D., van Os, J., Gross, V. A. (2007). Psychotic experiences in the general population: a twenty-year prospective community study. *Schizophrenia Research, 92,* 1–14.

Rozin, & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Rucker, D. D., McShane, B. B., & Preacher, K. J. (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology, 25*(4), 666-678.

Sánchez, P., Ojeda, N., Peña, J., Elizagárate, E., Yoller, A. B., Gutiérrez, M., & Ezcurra, J. (2009). Predictors of longitudinal changes in schizophrenia: the role of processing speed. *The Journal of clinical psychiatry*, *70*(6), 888-896.

Schnuerch, M., Nadarevic, L., & Rouder, J. N. (2021). The truth revisited: Bayesian analysis of individual differences in the truth effect. *Psychonomic Bulletin & Review*, *28*(3), 750–765. https://doi.org/10.3758/s13423-020-01814-8

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review, 2*(2), 87-99.

Sellen, J. L. (2005). Schizotypy and Conditional Reasoning. *Schizophrenia Bulletin, 31*(1), 105-116.

Shiloh, S., Salton, E., & Sharabi, D. (2002). Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. *Personality and Individual Differences*, *32*(3), 415-429.

Silva, R. R., Garcia-Marques, T., & Reber, R. (2017). The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth. *Consciousness and Cognition, 51,* 53-67.

Skurnik, I. W., Park, D. C., & Schwarz, N. (2000). Repeated warnings about false medical information can make it seem true: A paradoxical age difference. In *Eighth Cognitive Aging Conference, Atlanta, GA*.

So, S. H., Peters, E. R., Swendsen, J., Garety, P. A., & Kapur, S. (2013). Changes in delusions in the early phase of antipsychotic treatment – An experience sampling study. *Psychiatry Research, 215*(3), 568–573.

Spauwen, J., Krabbendam, L., Lieb, R., Wittchen, H. U., & van Os, J. (2003). Sex differences in psychosis: normal or pathological? *Schizophrenia Research, 62*(1-2), 45-49.

Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of Psychiatry and Neuroscience*, *35*(1), 7-17.

Spitzer, M., Braun, U., Maier, S., Hermle, L., & Maher, B. A. (1993). Indirect semantic priming in schizophrenic patients. *Schizophrenia Research*, *11*(1), 71-80.

Sprong, M., Schothorst, P., Vos, E., Hox, J., & Van Engeland, H. (2007). Theory of mind in schizophrenia. *British Journal of Psychiatry, 191*(1), 5-13.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society*, *11*, 3-13.

Startup, M., & Startup, S. (2005). On two kinds of delusion of reference. *Psychiatry Research*, *137*(1-2), 87-92.

Startup, M., Bucci, S., & Langdon, R. (2009). Delusions of reference: A new theoretical model. *Cognitive Neuropsychiatry, 14*(2), 110-126.

Stefanis, N. C., Hanssen, M., Smirnis, N. K., Avramoloulos, D. A., Evdokimidis, I. K., Stefanis, C. N., . . . Van Os, J. (2002). Evidence that three dimensions of psychosis have a distribution in the general population. *Psychological Medicine, 32*(2), 347-358.

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., ... & Corlett, P. R. (2018). The predictive coding account of psychosis. Biological psychiatry, 84(9), 634-643.

Sundar, A., Kardes, F. R., & Wright, S. A. (2015). The influence of repetitive health messages and sensitivity to fluency on the truth effect in advertising. *Journal of Advertising*, *44*(4), 375-387.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: a meta-analysis. *Psychological Bulletin*, *121*(3), 371-394.

Tabish. (2008). Complementary and Alternative Healthcare: Is it Evidence-based? *International Journal of Health Sciences, 2*(1), 5-9.

Tacikowski, P., Freiburghaus, T., & Ehrsson, H. H. (2017). Goal-directed processing of self-relevant information is associated with less cognitive interference than the processing of information about other people. *Journal of Experimental Social Psychology*, *68*, 93–100. https://doi.org/10.1016/j.jesp.2016.05.007.

Tandon, R., Nasrallah, H. A. & Keshavan, M. S. (2009). Schizophrenia, “just the facts” 4. Clinical features and conceptualization. *Schizophrenia Research*, *110*, 1-23.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events – the mobilization-minimization hypothesis. *Psychological Bulletin, 110,* 67–85.

Tomljenovic, H., Bubic, A., & Erceg, N. (2020). It just doesn’t feel right - the relevance of emotions and intuition for parental vaccine conspiracy beliefs and vaccination uptake. *Psychology & Health, 35*(5), 538–554. https://doi.org/10.1080/08870446.2019.1673894

Topolinski, S., & Strack, F. (2009). The analysis of intuition: Processing fluency and affect in judgements of semantic coherence. *Cognition and Emotion*, *23*(8), 1465-1503.

Tormala, Z. L. (2016). The role of certainty (and uncertainty) in attitudes and persuasion. *Current Opinion in Psychology*, *10*, 6-11.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 219-230.

Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 11-32). Psychology Press.

Unkelbach, C., & Greifeneder, R. (2018). Experiential fluency and declarative advice jointly inform judgments of truth. *Journal of Experimental Social Psychology*, *79*, 78-86.

Unkelbach, C., & Greifeneder, R. (Eds.). (2013). *The experience of thinking: How the fluency of mental processes influences cognition and behaviour*. Psychology Press.

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110-126.

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness & Cognition, 18*(1), 22-38.

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: explanations and implications. *Current Directions in Psychological Science, 28*(3), 247–253. https://doi.org/10.1177/0963721419827854

Van Os, J. (2003). Is there a continuum of psychotic experiences in the general population? *Epidemiology and Psychiatric Sciences*, *12*(4), 242-252.

Van Os, J., Hanssen, M., Bijl, R. V., & Ravelli, A. (2000). Strauss (1969) revisited: a psychosis continuum in the general population? *Schizophrenia research, 45*(1-2), 11-20.

Van Os, J., R, Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological Medicine, 39*(2), 179-195.

Van Os, J., Verdoux, H., & Hanssen, M. (1999). CAPE (Community Assessment of Psychic Experiences). *University of Maastricht, Maastricht, the Netherlands*.

Vartanian, O., Nakashima, A., Bouak, F., Smith, I., Baranski, J. V., & Cheung, B. (2012). Negative valence can evoke a liberal response bias in syllogistic reasoning. *Cognitive Processing, 14*(1), 89–98.

Vellani, V., Zheng, S., Ercelik, D., & Sharot, T. (2023). The illusory truth effect leads to the spread of misinformation. *Cognition*, *236*, 105421.

Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social forces*, *56*(2), 576-597.

Vogt, W. P. (2005). *Dictionary of statistics & methodology* (Vols. 1-0). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412983907

Vollema, M. G., & van den Bosch, R. J. (1995). The multidimensionality of schizotypy. *Schizophrenia bulletin*, *21*(1), 19-31.

Vollema, M. G., Sitskoorn, M. M., Appels, M. C. M., & Kahn, R. S. (2002). Does the Schizotypal Personality Questionnaire reflect the biological–genetic vulnerability to schizophrenia? *Schizophrenia Research, 54*(1-2), 39-45.

Vyse. (2013). *Believing in Magic.* Oxford University Press, Incorporated.

Walker, C. (1991). What did Jaspers really say? *Br J Psychiatry Suppl, 14,* 94-103.

Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science, 24*(3), 195–199. https://doi.org/10.1177/0963721414561766.

Ward, T., & Garety, P. A. (2019). Fast and slow thinking in distressing delusions: A review of the literature and implications for targeted therapy. *Schizophr Res, 203*, 80-87.

Westbrook, D., Kennerley, H., & Kirk, J. (2011). *An introduction to cognitive behaviour therapy: Skills and applications*. Sage.

Westerman, D. L., Lanska, M., & Olds, J. M. (2015). The effect of processing fluency on impressions of familiarity and liking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 426.

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235.

Whittlesea, B. W., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologia, 98*(2-3)*,* 141-165.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*(6), 716-732.

Williams, E. B. (1964). Deductive reasoning in schizophrenia. *Journal of Abnormal Psychology, 69,* 47–61.

Wiswede, D., Koranyi, N., Muller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: behavioral and ERP indicators of truth evaluation processes. *Soc Cogn Affect Neurosci, 8*(6), 647-653.

Witteman, C., Van Den Bercken, J., Claes, L., & Godoy, A. (2009). Assessing rational and intuitive thinking styles. *European Journal of Psychological Assessment*, *25*(1), 39–47. https://doi.org/10.1027/1015-5759.25.1.39

Wolfradt, U., Oubaid, V., Straube, E. R., Bischoff, N., & Mischo, J. (1999). Thinking styles, schizotypal traits and anomalous experiences. *Personality and Individual Differences*, *27*(5), 821-830.

Woodward, T. S., Moritz, S., Cuttler, C., & Whitman, J. C. (2006). The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions in schizophrenia. *Journal of Clinical and Experimental Neuropsychology*, *28*(4), 605-617.

World Health Organisation. (2022, January 24). *Schizophrenia.* WHO. https://www.who.int/news-room/fact-sheets/detail/schizophrenia

Xu, L., Zhang, M., Wang, S., Wei, Y., Cui, H., Qian, Z., ... & Wang, J. (2021). Relationship between cognitive and clinical insight at different durations of untreated attenuated psychotic symptoms in high-risk individuals. *Frontiers in Psychiatry*, *12*, 753130.

Zander, T., Fernandez Cruz, A. L., Winkelmann, M. P., & Volz, K. G. (2017). Scrutinizing the Emotional Nature of Intuitive Coherence Judgments. *Journal of Behavioral Decision Making, 30*(3), 693–707.

Zimmermann, G., Favrod, J., Trieu, V. H., & Pomini, V. (2005). The effect of cognitive behavioral treatment on the positive symptoms of schizophrenia spectrum disorders: a meta-analysis. *Schizophrenia Research*, *77*(1), 1-9.

APPENDIX A

Experiment 1 and 2 Materials

Delusion-relevant and neutral statements

Contamination

1. (F) Exposure to radiation on a long-haul flight from New York to Tokyo is approximately 5% of the dose at which genetic damage can occur.
2. (T) The WHO estimates that 865,000 people worldwide die every year as a result of poor air quality in cities alone.
3. (T) Although diluted during the vaccine manufacturing process, residual quantities of formaldehyde may be found in some current vaccines.
4. (T) Microorganisms inhabiting the human body outnumber human cells by 10 to 1.
5. (T) Approximately 1 to 3 kg of a 100 kg human adult is bacteria (that is, 1 – 3% of body mass).
6. (T) On the face of every human being live hundreds to thousands of microscopic mites.
7. (F) Almonds contain trace amounts of arsenic.
8. (F) Infant formula contains high levels of fluoride when mixed with fluoridated water.
9. (F) Placing hot food immediately in a refrigerator is unsafe.
10. (F) Food poisoning caused by parasites is not contagious.

Persecution

11. (F) The resolution of Earth observation satellites is so high that it is possible to read car license plates from a height of 300 km.
12. (T) Within the framework of the MK Ultra project, the CIA investigated, among other things, the possibilities of mind control through LSD.
13. (F) Between 2018-2019, 7,452 telecommunication interception warrants were issued in Australia.
14. (F) Australian state and local police agencies can access an individual's telecommunications data without a warrant.
15. (F) The Australian Secret Intelligence Service (ASIS) was formed in 1952, but it remained a secret even within the Government until 1979.
16. (T) In the process of spying on foreigners, the USA's National Security Agency incidentally collects large volumes of Americans' communications, including conversations, emails, and photographs.
17. (T) Federal authorities such as the FBI can search American's personal data without a warrant.

18. (T) Facebook has one of the most powerful and accurate AI-based facial recognition systems on the planet, which learns from every photograph that is tagged.
19. (T) Facebook stores biometric data from people who no longer have, or have never had, a Facebook account.
20. (F) In 2018/19, there were 254 police-related fatalities in England and Wales.

Conspiracy

21. (T) With the knowledge and tolerance of the CIA, tonnes of drugs were smuggled into the USA in the 1980s.
22. (T) Until 1977 the codes for the security of all US nuclear missiles stationed in Europe were constantly set to "00000000" (8 times zero).
23. (T) One of the justifications for the 1991 Iraq war was the fabricated claim that Iraqi soldiers in Kuwait killed babies in their incubators.
24. (T) Banjawarn Station, a remote cattle station in Western Australia, was once owned by a Japanese doomsday cult that was implicated in criminal acts of violence and terrorism.
25. (T) During Prohibition in the United Sates, the government literally poisoned alcohol.
26. (F) Water condensation trails (“contrails”) from aircraft contain a combination of water, aluminium, strontium and barium.
27. (F) The Bermuda Triangle is one of only two places on Earth where a compass points to magnetic north rather than true north.
28. (T) In the early 90s, Pepsi owned 17 submarines after exchanging soda for military equipment with the Soviet Union.
29. (F) In the mid-1990s, a black-and-white film emerged, depicting a team of scientists and doctors performing an alien autopsy at Area 51.
30. (F) There is still no known explanation for one of the world's most complex and spectacular crop circles, which appeared near Stonehenge, England, in July 1996.

Supernatural

31. (T) Tetraphobia is an avoidance or fear of the number four. It is a commonly held superstition in many Asian countries, including China, Vietnam, Korea, and Japan.
32. (T) In Russia, celebrating or even congratulating someone on a birthday before the actual day is considered to bring bad luck.
33. (T) In Mexico, placing two mirrors opposite each other is believed to open a doorway for the devil.
34. (T) Getting a haircut on a Wednesday is considered bad luck in India.
35. (T) Whistling indoors is forbidden in Lithuania, where people believe the noise can summon demons.

36. (F) There is a superstition in Spain stating that one who says “cheers” with water is actually wishing death upon those with whom they are drinking.
37. (F) Sleeping with one’s head to the South is considered bad luck in Japan.
38. (F) In Vietnam, an itchy right hand means one will come into some money but an itch on one’s left means a loss of money.
39. (F) In France, stepping in dog poo with one’s right foot is considered good luck.
40. (T) Syria banned Yo-Yos in 1933 over fear that they would cause a drought.

Complementary and alternative medicinal practices

41. (T) Maggot debridement therapy (MDT) is an effective method of wound treatment in which sterile or disinfected larvae of the green bottle fly are used to feed on tissue, pus, and metabolic wastes in order to promote healing.
42. (F) Homeopathy is an effective alternative to antibiotics in infectious diseases, producing no toxic side effects and bringing about rapid recovery.
43. (F) A recent meta-analysis investigating the effects of acupuncture on hypertension found the technique to be effective as a stand-alone treatment.
44. (T) Recent scientific findings support heartrate variability biofeedback (which involves feedback of one’s heartrate through a device) as a potential intervention to improve motor function in athletes.
45. (T) With the possible exception of back pain, chiropractic spinal manipulation has not been shown to be effective for any medical condition.
46. (F) Coffee enemas have been shown to cure mild autism.
47. (F) Reiki helps with symptoms of asthma, chronic fatigue, menopausal symptoms and arthritis.
48. (F) The benefits of reflexology include its ability to stimulate nerve function, increase energy, boost circulation and eliminate toxins from the body.
49. (T) Studies have shown that the practice of yoga can decrease the secretion of cortisol – a stress hormone – in the body.

Neutral statements

Science

50. (T) The science of agriculture is called agronomy.
51. (T) The goldfish sees the widest colour spectrum of all living creatures.
52. (T) The marine animal the ‘mantis shrimp’ is neither a mantis nor a shrimp.
53. (T) The trachea of a human being is about 12 cm long.
54. (F) A python can live without food for up to 18 months.

55. (F) Giraffes do not have vocal cords.
56. (T) The smallest known living mammal is the bumblebee bat.
57. (F) The moon fits about sixty times into the globe.
58. (F) Ivy is the fastest growing plant.
59. (F) The tongue is the strongest muscle in the human body.

History

60. (T) Finland was the first European country to introduce the right to vote for women.
61. (T) The plane that dropped the atomic bomb over Nagasaki was called "Bock's Car".
62. (F) The body of Alexander the Great was preserved in a barrel of rum.
63. (T) The 1918 Spanish influenza did not originate in Spain.
64. (T) The first living creature in space was a dog named Laika.
65. (T) The infamous pirate Blackbeard's beloved ship was called Queen Anne's Revenge.
66. (F) The coldest day ever recorded was -85ºC (-121ºF).
67. (F) In the 1800s, the Olympic Games awarded medals for the arts.
68. (F) Cleopatra was the descendant of a Roman general.
69. (F) In the 13th Century, Pope Gregory IV declared a war on rats.

Geography

70. (T) On every continent there is a city called Rome.
71. (F) There are 3 natural rivers in Saudi Arabia.
72. (T) The Danube flows through 10 countries.
73. (T) Of all countries, Monaco has the highest population density.
74. (F) Bolivia is the smallest landlocked country in South America.
75. (F) Zambia was formerly called Southern Rhodesia.
76. (F) The highest altitude city in the world lies in Ecuador.
77. (T) Standing shoulder to shoulder, the entire world's population could fit within Los Angeles.
78. (F) The Canary Islands are named after cats, not birds.

Pop culture

79. (F) Meg Ryan was a journalist before her Hollywood breakthrough.
80. (T) Ben Affleck's father was Robert Downey Jr's drug and alcohol use counsellor.
81. (T) Tom Cruise is dyslexic.

82. (F) Donald Trump's middle name is James.
83. (F) Ellen DeGeneres is distantly related to Meghan Markle, Duchess of Sussex.
84. (F) Kourtney Kardashian suffered a traumatic brain injury in 2001 after a car accident.
85. (T) Justin Timberlake's mom became Ryan Gosling's legal guardian when they were 12.
86. (T) Woody Harrelson's father, Charles Harrelson, was a hitman who was charged with multiple murders.
87. (T) Breaking Bad's Bryan Cranston was once a suspect in a murder investigation.
88. (F) Joe Exotic A.K.A. The Tiger King's real name is Joseph Andrew McArthur-Passage.
89. (T) KFC once developed a chicken-flavoured nail polish.

General trivia

90. (T) The Hawaiian alphabet consists of only 12 letters.
91. (F) In Nepal, Mount Everest is called "Holy Mountain".
92. (F) Male canaries cannot sing.
93. (F) Humans and monkeys are the only creatures that cry for emotional reasons.
94. (F) A person who is afraid of work suffers from laboratory apathy.
95. (T) Minerva is the Roman goddess of wisdom.
96. (F) Before there were mercury thermometers, thermometers were filled with distilled water.
97. (T) Mahatma Gandhi married at the age of 13 years.
98. (T) Myanmar, Liberia and the US are the only three countries in the world that have not adopted the metric system.

* PLEASE NOTE: To the best of our knowledge, these are the correct responses to the trivia questions. However, we acknowledge the possibility that there may be some inaccuracies in our current understanding.

APPENDIX B

Experiment 5 Materials

Delusion-relevant and neutral statements

Complementary and alternative medicinal practices

1. (T) Emerging research supports the use of sulforaphane, a supplement derived from cruciferous vegetables, in reducing some symptoms of autism.

2. (F) Scientific studies have shown Calcarea carb as effective in preventing recurrent upper respiratory tract infections in children.
3. (F) A recent meta-analysis investigating the effects of acupuncture on hypertension found the technique to be effective as a stand-alone treatment.
4. (T) Recent studies indicate that electroacupuncture may prevent or reduce chronic stress by blocking stress-induced hormone elevations in the nervous system.
5. (T) Alternating bilateral sensory stimulation (ABS) has been shown effective in the treatment of traumatic stress by suppressing activation of the amygdala in the brain.
6. (F)Coffee enemas have been shown to cure mild autism.
7. (F) Reiki helps with symptoms of asthma, chronic fatigue, menopausal symptoms and arthritis.
8. (F)The benefits of reflexology include its ability to stimulate nerve function, increase energy, boost circulation and eliminate toxins from the body.

Conspiracy

9. (T) In 1964, the US set off nuclear bombs under Mississippi.
10. (T) Until 1977 the codes for the security of all US nuclear missiles stationed in Europe were constantly set to "00000000" (8 times zero).
11. (F) Water condensation trails ("contrails") from aircraft contain a combination of water, aluminium, strontium and barium.
12. (T) One of the justifications for the 1991 Iraq war was the fabricated claim that Iraqi soldiers in Kuwait killed babies in their incubators.
13. (T) Banjawarn Station, a remote cattle station in Western Australia, was once owned by a Japanese doomsday cult that was implicated in criminal acts of violence and terrorism.
14. (F) The Bermuda Triangle is one of only two places on Earth where a compass points to magnetic north rather than true north.
15. (T) In the early 90s, Pepsi owned 17 submarines after exchanging soda for military equipment with the Soviet Union.
16. (T) During Prohibition in the United States, the government literally poisoned alcohol.
17. (F) In the mid-1990s, a black-and-white film emerged, depicting a team of scientists and doctors performing an alien autopsy at Area 51.

Contamination

18. (F) Exposure to radiation on a long-haul flight from New York to Tokyo is approximately 5% of the dose at which genetic damage can occur.
19. (T) If there are fewer than 100 CFU (colony forming units) of Staphylococcus aureus (golden staph) per gram in a sample of food, the product is considered safe.

20. (T) Although diluted during the vaccine manufacturing process, residual quantities of formaldehyde may be found in some current vaccines.
21. (T) Approximately 1 to 3 kg of a 100 kg human adult is bacteria (that is, 1 - 3% of body mass).
22. (F) Infant formula contains high levels of fluoride when mixed with fluoridated water.
23. (F) Food poisoning caused by parasites is not contagious.
24. (T) Hot water is no more effective than cold water when cleaning away harmful bacteria from skin and other surfaces.
25. (F) The resolution of Earth observation satellites is so high that it is possible to read car license plates from a height of 300 km.

Persecution

26. (T) The last English woman tried for witchcraft was convicted in 1944.
27. (F) Between 2018-2019, 7,452 telecommunication interception warrants were issued in Australia.
28. (F) In 2018/19, there were 254 police-related fatalities in England and Wales.
29. (F) Australian state and local police agencies can access an individual's telecommunications data without a warrant.
30. (F) The Australian Secret Intelligence Service (ASIS) was formed in 1952, but it remained a secret even within the Government until 1979.
31. (T) Only twenty percent of the world's CCTV is in the United Kingdom.
32. (T) Regular air traffic control radars can very often see the most advanced stealth bombers.
33. (T) The US Defense's I.C.E. program aims to control the physical properties of ice crystals to protect military assets and personnel in extreme cold environments.

Supernatural

34. (T) In Russia, celebrating or even congratulating someone on a birthday before the actual day is considered to bring bad luck.
35. (F) There is a superstition in Spain stating that one who says "cheers" with water is actually wishing death upon those with whom they are drinking.
36. (F) Sleeping with one's head to the South is considered bad luck in Japan.
37. (T) In Mexico, placing two mirrors opposite each other is believed to open a doorway for the devil.
38. (F) Getting a haircut on a Wednesday is considered bad luck in India.
39. (F) In Vietnam, an itchy right hand means one will come into some money but an itch on one's left means a loss of money.

40. (T) Whistling indoors is forbidden in Lithuania, where people believe the noise can summon demons.
41. (F) In France, stepping in dog poo with one's right foot is considered good luck.
42. (T) Syria banned Yo-Yos in 1933 over fear that they would cause a drought.

Neutral statements

Geography

43. (T) On every continent there is a city called Rome.
44. (F) There are 3 natural rivers in Saudi Arabia.
45. (T) Although Mt Everest is the highest mountain on Earth, Mount Chimborazo is actually the tallest.
46. (T) Of all countries, Monaco has the highest population density.
47. (F) Bolivia is the smallest landlocked country in South America.
48. (F) Zambia was formerly called Southern Rhodesia.
49. (F) The highest altitude city in the world lies in Ecuador.
50. (T) Standing shoulder to shoulder, the entire world's population could fit within Los Angeles.
51. (T) The Canary Islands are named after dogs, not birds.

History

52. (T) The infamous pirate Blackbeard was only a pirate for two years.
53. (F) The coldest day ever recorded was -85°C (-121°F).
54. (F) In the 1800s, the Olympic Games awarded medals for the arts.
55. (T) When the Pyramids were built, woolly mammoths still roamed the Earth.
56. (T) The plane that dropped the atomic bomb over Nagasaki was called "Bock's Car".
57. (T) The earliest documented case of the Spanish flu was in Kansas, United States, in 1918.
58. (F) Cleopatra was the descendant of a Roman general.
59. (F) In the 13th Century, Pope Gregory IV declared a war on rats.

Pop culture

60. (T) Woody Harrelson's father, Charles Harrelson, was a hitman who was charged with multiple murders.
61. (T) Breaking Bad's Bryan Cranston was once a suspect in a murder investigation.
62. (F) Donald Trump's middle name is James.

63. (F) Ellen DeGeneres is distantly related to Meghan Markle, Duchess of Sussex.
64. (F) Kourtney Kardashian suffered a traumatic brain injury in 2001 after a car accident.
65. (T) Joe Exotic A.K.A. The Tiger King's real name is Joseph Allen Maldonado-Passage.
66. (F) Meg Ryan was a journalist before her Hollywood breakthrough.
67. (T) Ben Affleck's father was Robert Downey Jr's drug and alcohol use counsellor.

Science

68. (T) The science of agriculture is called agronomy.
69. (T) The goldfish sees the widest colour spectrum of all living creatures.
70. (F) A python can live without food for up to 18 months.
71. (T) The marine animal the 'mantis shrimp' is neither a mantis nor a shrimp.
72. (F) Giraffes do not have vocal cords.
73. (F) The tongue is the strongest muscle in the human body.
74. (T) The smallest known living mammal is the bumblebee bat.
75. (F) The moon fits about sixty times into the globe.
76. (F) Ivy is the fastest growing plant.

General trivia

77. (T) The Hawaiian alphabet consists of only 12 letters.
78. (F) In Nepal, Mount Everest is called "Holy Mountain".
79. (F) Male canaries cannot sing.
80. (F) Humans and monkeys are the only creatures that cry for emotional reasons.
81. (T) Chewbacca’s voice was created by combining the sounds of a bear, a walrus, a lion and a badger.
82. (F) Before there were mercury thermometers, thermometers were filled with distilled water.
83. (T) Mahatma Gandhi married at the age of 13 years.
84. (T) Einstein’s last words were spoken in German to a nurse who didn’t speak German and are lost for ever.

* PLEASE NOTE: To the best of our knowledge, these are the correct responses to the trivia questions. However, we acknowledge the possibility that there may be some inaccuracies in our current understanding.

APPENDIX C

Schizotypy and Psychosis-Proneness Measures

The Oxford-Liverpool Inventory of Feelings and Experiences (Short Version)

Unusual experiences (12 items)

1. When in the dark do you often see shapes and forms even though there is nothing there?
2. Are your thoughts sometimes so strong that you can almost hear them?
3. Have you ever thought that you had special, almost magical powers?
4. Have you sometimes sensed an evil presence around you, even though you could not see it?
5. Do you think that you could learn to read other's minds if you wanted to?
6. When you look in the mirror does your face sometimes seem quite different from usual?
7. Do ideas and insights sometimes come to you so fast that you cannot express them all?
8. Can some people make you aware of them just by thinking about you?
9. Does a passing thought ever seem so real it frightens you?
10. Do you feel that your accidents are caused by mysterious forces?
11. Do you ever have a sense of vague danger or sudden dread for reasons that you do not understand?
12. Does your sense of smell sometimes become unusually strong?

Cognitive disorganisation (12 items)

13. Are you easily confused if too much happens at the same time?
14. Do you frequently have difficulty in starting to do things?
15. Are you a person whose mood goes up and down easily?
16. Do you dread going into a room by yourself where other people have already gathered and are talking?
17. Do you find it difficult to keep interested in the same thing for a long time?
18. Do you often have difficulties in controlling your thoughts?
19. Are you easily distracted from work by daydreams?

20. Do you ever feel that your speech is difficult to understand because the words are all mixed up and don't make sense?
21. Are you easily distracted when you read or talk to someone?
22. Is it hard for you to make decisions?
23. When in a crowded room, do you often have difficulty in following a conversation?

Introvertive anhedonia (10 items)

24. Are there very few things that you have ever enjoyed doing?
25. Are you much too independent to get involved with other people?
26. Do you love having your back massaged? [a]
27. Do you find the bright lights of a city exciting to look at? [a]
28. Do you feel very close to your friends? [a]
29. Has dancing or the idea of it always seemed dull to you?
30. Do you like mixing with people? [a]
31. Is trying new foods something you have always enjoyed? [a]
32. Have you often felt uncomfortable when your friends touch you?
33. Do you prefer watching television to going out with people?

Impulsive nonconformity (10 items)

34. Do you consider yourself to be pretty much an average sort of person? [a]
35. Would you like other people to be afraid of you?
36. Do you often feel the impulse to spend money which you know you can't afford?
37. Are you usually in an average kind of mood, not too high and not too low? [a]
38. Do you at times have an urge to do something harmful or shocking?
39. Do you stop to think things over before doing anything? [a]
40. Do you often overindulge in alcohol or food?
41. Do you ever have the urge to break or smash things?
42. Have you ever felt the urge to injure yourself?
43. Do you often feel like doing the opposite of what other people suggest even though you know they are right?

[a] Score 1 for no, 0 for yes

The Community Assessment of Psychic Experiences (Short Version)

1. Do you ever feel sad?
2. Do you ever feel as if people seem to drop hints about you or say things with a double meaning?
3. Do you ever feel that you are not a very animated person?
4. Do you ever feel that you are not much of a talker when you are conversing with other people?
5. Do you ever feel as if things in magazines or on TV were written especially for you?
6. Do you ever feel as if some people are not what they seem to be?
7. Do you ever feel as if you are being persecuted in some way?
8. Do you ever feel that you experience few or no emotions at important events?
9. Do you ever feel pessimistic about everything?
10. Do you ever feel as if there is a conspiracy against you?
11. Do you ever feel as if you are destined to be someone very important?
12. Do you ever feel as if there is no future for you?
13. Do you ever feel that you are a very special or unusual person?
14. Do you ever feel as if you do not want to live anymore?
15. Do you ever think that people can communicate telepathically?
16. Do you ever feel that you have no interest to be with other people?
17. Do you ever feel as if electrical devices such as computers can influence the way you think?
18. Do you ever feel that you are lacking in motivation to do things?
19. Do you ever cry about nothing?
20. Do you believe in the power of witchcraft, voodoo or the occult?
21. Do you ever feel that you are lacking in energy?
22. Do you ever feel that people look at you oddly because of your appearance?
23. Do you ever feel that your mind is empty?
24. Do you ever feel as if the thoughts in your head are being taken away from you?
25. Do you ever feel that you are spending all your days doing nothing?
26. Do you ever feel as if the thoughts in your head are not your own?
27. Do you ever feel that your feelings are lacking in intensity?

28. Have your thoughts ever been so vivid that you were worried other people would hear them?
29. Do you ever feel that you are lacking in spontaneity?
30. Do you ever hear your own thoughts being echoed back to you?
31. Do you ever feel as if you are under the control of some force or power other than yourself?
32. Do you ever feel that your emotions are blunted?
33. Do you ever hear voices when you are alone?
34. Do you ever hear voices talking to each other when you are alone?
35. Do you ever feel that you are neglecting your appearance or personal hygiene?
36. Do you ever feel that you can never get things done?
37. Do you ever feel that you have only few hobbies or interests?
38. Do you ever feel guilty?
39. Do you ever feel like a failure?
40. Do you ever feel tense?
41. Do you ever feel as if a double has taken the place of a family member, friend or acquaintance?
42. Do you ever see objects, people or animals that other people cannot see?

Note. Items in each trait dimension and response options are as below.
The positive dimension consists of items 2, 5, 6, 7, 10, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 31, 33, 34, 41, and 42.
The depressive dimension consists of items 1, 9, 12, 14, 19, 38, 39, and 40.
The negative dimension consists of items 3, 4, 8, 16, 18, 21, 23, 25, 27, 29, 32, 35, 36, and 37.
Available responses for each item: Never, Sometimes, Often, and Nearly always.
All responses, with the exception of "Never", lead to a subsequent question: *Please indicate how distressed you are by this experience.* Available responses to this question: Not distressed, A bit distressed, Quite distressed, and Very distressed.

APPENDIX D

Experiment 6 Materials

Referential item descriptive words

List 1	List 2	List 3	List 4	List 5	List 6
agreeable	academic	aggressive	ambitious	analytic	alternative
angry	active	authoritative	appreciative	anxious	antisocial
artistic	altruistic	beautiful	athletic	cynical	attractive
awkward	amiable	brave	conservative	emotional	cheeky
bold	complex	conscientious	content	fortunate	childish
diplomatic	cool	curious	crude	funny	clever
gloomy	different	difficult	fashionable	generous	competitive
gracious	facetious	economical	hilarious	just	courteous
individual	friendly	elusive	honest	nervous	critical
intellectual	imaginative	enthusiastic	humble	noble	dramatic
intense	lazy	happy	impressive	philosophical	earnest
jolly	little	impatient	lovely	polite	efficient
kind	logical	interesting	mellow	popular	elegant
lively	mysterious	modest	morbid	practical	gentle
negative	neat	peculiar	open	proud	good
noisy	patient	persuasive	responsible	radical	intelligent
pedantic	poetic	quiet	rude	respectable	jealous
pleasant	private	realistic	short	serious	loud
positive	reasonable	sad	smart	small	lucky
sensible	romantic	sceptical	special	tough	nice
sensitive	sentimental	shy	strange	unlucky	orderly
squeamish	sociable	tolerant	strong	vivacious	original
tender	temperamental	unique	tall	volatile	progressive
trustworthy	wise	unpredictable	unhappy	weird	silly