AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Abnormal Psychology

© 2021 American Psychological Association
ISSN: 0021-843X

2021, Vol. 130, No. 3, 273–285
https://doi.org/10.1037/abn0000664

Paranoia Is Associated With Impaired Novelty Detection and Overconfidence in Recognition Memory Judgments

William N. Koller and Tyrone D. Cannon
Department of Psychology, Yale University

False recognition, or the mis-categorization of a "new" stimulus as "old," might support fixed false beliefs by blocking new learning or otherwise contributing to internal representations of the world that are at odds with reality. However, the mechanisms through which false recognition is facilitated among paranoid individuals remain unclear. We examined 2 phenomena that may contribute to this effect: an overreliance on fluency-based processes during recognition, manifesting as a lower threshold for judging items as recently studied, and a propensity to require less information to come to a highly confident judgment. The former would be expected to be particularly pronounced among items that are generally familiar, as opposed to completely novel. Here, we manipulated familiarity in a recognition memory paradigm by using stimuli that varied in their rate of extraexperimental exposure (i.e., real words vs. pseudowords). Further, to determine whether paranoia was associated with a tendency to differentially misallocate confidence to errors, we calculated a hierarchical Bayesian estimate of metacognitive sensitivity (meta-d') in addition to the more classic d'. In line with our hypotheses, paranoia was associated with an increased rate of false alarm errors, differentially so for familiar versus unfamiliar stimuli, suggesting that a context-agnostic, familiarity-based memory system might underlie observed memory distortions. What's more, paranoia was associated with heightened confidence on error trials and reduced metacognitive sensitivity. These findings highlight 2 distinct deficits—in both novelty detection and metacognitive monitoring—that contribute to false recognition judgments, offering targets for cognitive interventions to reduce memory distortion among paranoid individuals.

General Scientific Summary

This study showed that people high in paranoid thinking were more likely to falsely recognize a new item as having been previously presented. We identified two processes that contributed to this effect: (a) an overreliance on nonspecific feelings of familiarity when making memory judgments and (b) overconfidence in those judgments, potentially related to a reduced ability to question one's own mistakes. These findings offer some insight into the ways in which new information might be disregarded by paranoid individuals and how this might contribute to models of the world that are out-of-sync with reality and resistant to correction.

Keywords: paranoia, memory, novelty detection, overconfidence, metacognition

Supplemental materials: https://doi.org/10.1037/abn0000664.supp

Disturbances of cognition are a core feature of schizophrenia and related disorders, with memory impairments counting among the most prominent areas of deficit (Aleman et al., 1999; Guo et al., 2019). Given that the encoding of new information and the retrieval of items stored in memory play key roles in learning and belief revision, aberrancies in memory-related processes may participate in the formation and maintenance of false beliefs. In their most extreme form, such beliefs are known as delusions—a prominent positive symptom of schizophrenia (Butler & Braff, 1991). However, many individuals without a diagnosis of schizophrenia also espouse less extreme forms of such false beliefs, suggesting that these symptoms may have some level of dimensionality. Dimensional models of schizophrenia suggest that its features extend across a spectrum of "schizotypy" made up of individuals who may never receive a diagnosis of schizophrenia but who nevertheless exhibit many of its cardinal symptoms, albeit to a less extreme extent (Lenzenweger, 2010). Based on this model, schizotypy may prove to be a convenient translational framework for exploring in the general population processes that might facilitate

This article was published Online First January 25, 2021.
William N. Koller https://orcid.org/0000-0001-5117-2847
Tyrone D. Cannon https://orcid.org/0000-0002-5632-3154
We have no conflict of interest to disclose. This study's protocol (2000026576) was deemed exempt by the Yale University IRB. Data in this article have not been previously published.

Correspondence concerning this article should be addressed to William N. Koller, Department of Psychology, Yale University, P.O. Box 208205, 2 Hillhouse Avenue, New Haven, CT 06520-8205, United States. Email: william.koller@yale.edu

false belief in its most extreme forms. However, the extent to which the memory impairments characteristic of schizophrenia A) extend to individuals in the general population and B) co-occur with common symptoms of positive schizotypy such as paranoia remains unclear.

Recent research has begun to expand our knowledge of the associations between memory impairment and schizotypy. For example, Sahakyan and Kwapil (2019) found that recognition memory deficits extended to individuals with symptoms of schizotypy drawn from the general population. Further, they found differential patterns of deficits in relation to negative versus positive symptoms of schizotypy—the former group being characterized by a decreased rate of hits despite a stable false alarm rate; the latter by an increased rate of false alarms despite a stable hit rate. In this study, we build upon this work by focusing on two specific processes—a deficit in novelty detection and overconfidence in recognition memory judgments—that might contribute to elevated false recognition rates among individuals endorsing experiences of paranoia, a relatively common area of delusion content among the general population.

False Recognition and Delusionality

Memory supports and shapes our internal models of our environment— both our recollections of how it was and our predictions of how it will be. Delusional belief can be broadly understood as a mismatch between one's internal model and the external world, in which the subjective experiences, expectations, and convictions of an individual are out of keeping with consensual reality. Thus, memory distortion may represent one mechanism through which fixed false belief is perpetuated—a deficit in one's ability to encode and contrast events in memory may contribute to rigid internal representations that are not reflective of one's external world.

False recognition, or the mis-categorization of a "new" stimulus as "old," may support delusional beliefs such as paranoia. This could be facilitated by its contribution to cognitive biases that frequently co-occur with delusionality. For example, an overly liberal threshold for judging a stimulus as familiar may interfere with one's ability to notice and encode novel information that may be inconsistent with a strongly held belief. In this way, novelty detection deficits might support a bias against disconfirmatory evidence, a well-established feature of delusional ideation (Bronstein & Cannon, 2017; Woodward et al., 2007; Woodward et al., 2006). More generally, given the importance of novelty detection to the acquisition of new learning (e.g., Tulving et al., 1996), we believe that characterizing novelty detection deficits among paranoid individuals may contribute to the emerging associative account of delusions, which holds associative learning as a key process whose dysfunction supports delusional beliefs (Corlett et al., 2007). In short, we hypothesize that novelty detection deficits may impinge upon one's ability to engage in adaptive and flexible learning about one's environment, thereby supporting false beliefs that are resistant to updating.

While theoretical links can be drawn between novelty detection deficits and general delusionality, the extent of the association between novelty detection and paranoia remains to be elucidated. To this end, the present study aims to evaluate the extent to which novelty detection deficits are associated with self-reported persecutory ideation. Persecutory ideation is an area of delusion content that is present in over 70% of first-episode psychosis patients (Freeman, 2016) and is relatively prevalent in the general population in less extreme forms (nearly 20% of a nonclinical sample of over 7000 individuals; Freeman et al., 2011). For this reason, paranoia is a natural target for studies seeking to understand delusion-relevant processes in nonclinically ascertained samples. Further, the impact of paranoia among the general public is not insignificant, with concomitants that range from poor social functioning to heightened suicidal ideation (Freeman et al., 2011) – understanding the cognitive processes that co-occur with this symptom is important for the development of interventions that reduce feelings of paranoia and suspiciousness among the general population as well as in clinical samples.

Familiarity-Based Memory

An overreliance on fluency-based processes is likely to represent a core component of false recognition memories. One classical method of generating "false memories" comes from Deese (1959) and Roediger and McDermott (1995). The so-called "DRM" effect manifests as an increased propensity to falsely recall having seen a lure (i.e., a newly presented item that was not studied during an encoding phase) if it is highly semantically related to items on the study list (e.g., study list = "hospital," "medicine," "nurse"; lure = "doctor"). Theoretically, the fluency with which this critical lure is processed increases its odds of being falsely judged as having been previously encountered. In other words, the semantic relatedness of lures to studied items creates a sense of familiarity that is sufficient for an item to be falsely recognized as "old" despite its novelty—an effect that is present to varying degrees in people from the general population.

In studies using the "Remember-Know" paradigm, individuals on the psychosis spectrum tend to display an overreliance on familiarity-based processes to inform their recollective judgments, even in the absence of semantic associations between targets and lures (Achim & Lepage, 2005). In this paradigm, a "remember" response is characterized as an instance of conscious recollection in which one clearly recollects the item in question and can recall specific features of this item. This is contrasted with a "know" response, in which one's memory judgment is based on a gist-based sense of familiarity; accordingly, "know" responses are characterized by a reduced ability to recall additional details about the features of the item or the context in which it appeared. A meta-analysis of remember-know literature supports the notion that patients with schizophrenia show greater impairment in conscious recollection (i.e., "remember") relative to familiarity-based assessments (i.e., "know"; Libby et al., 2013). This reliance on familiarity-based memory may be related to impairments in novelty detection. For example, van Erp et al. (2008) reported a lower "old-new" criterion based on a unidimensional model of memory strength, as well as an increased proportion of "know" responses relative to "remember" responses, among a sample of patients with schizophrenia.

Familiarity-based memory loads differentially onto neural circuitry within the medial temporal lobe (MTL), with canonical findings associating this type of assessment with the perirhinal cortex (Dew & Cabeza, 2013), compared to conscious recollection, which relies more heavily on the hippocampus itself (Diana et al., 2007). Imaging studies have revealed abnormal patterns of activation in the hippocampus during both encoding and retrieval among patients with schizophrenia (e.g., Jessen et al., 2003; Weiss et al., 2004) as well as structural abnormalities of the hippocampus among individuals at clinical high risk who convert to schizophrenia (Provenzano et al., 2020). Further, Weiss et al. (2004) observed functional abnormalities within the hippocampus specifically in the context of novel word detection. Schott et al. (2015) and Tamminga et al. (2012) also found evidence for disrupted novelty processing in the MTL in samples with schizophrenia. In the former, a dissociation between novelty-related hippocampal activity and recognition memory performance was observed among patients with paranoid schizophrenia but not within a healthy control group. In the latter, blunted novelty-related activity in the hippocampus and parahippocampal cortex was observed among nonmedicated (but not medicated) patients. Together, these studies lend support to a link between MTL dysfunction—in the hippocampus, in particular—and impaired novelty detection among individuals on the psychosis spectrum; reliance on familiarity-based recognition judgments may help to compensate for such deficits.

Given that a reliance on fluency-based heuristics makes it difficult to tag particular instances of a stimulus to a particular learning context (Dudukovic & Knowlton, 2006), a familiarity-based memory tends to be impoverished in terms of contextual detail. Accordingly, such a system may be particularly prone to errors if it is tasked with detecting stimuli that are contextually, but not absolutely, novel. In other words, one would expect an individual who relies on familiarity-based memory processes to show an elevated false alarm rate for items that are accompanied by a *general sense of familiarity* versus items that have *never been encountered* and for which a sense of familiarity (or a lack thereof) could be used as a more meaningful proxy for conscious recollection. The ability to detect mismatches in novelty between item and context—a key function in detecting contextual but not absolute novelty—may also be hippocampally mediated (Thakral et al., 2015), and thus may be impaired if the hippocampus is not functioning in a normative manner. Ragland et al. (2015) suggest that this might be the case among patients with schizophrenia, who displayed specific impairments in relational (vs. item-specific) encoding in association with blunted hippocampal activation.

In summary, past research suggests that, in general, individuals with schizophrenia may struggle to identify "new" items in their environment due to a memory system characterized by impoverished episodic detail. Decreased recollective specificity for target stimuli may increase the odds that a lure, even if only loosely matching a studied item, is incorrectly accepted as "old". Furthermore, as such a system tends to be accompanied by an overreliance on fluency-based cues and a lack of context-specificity, these individuals may experience a particularly elevated rate of false recognition among items that are experienced as generally familiar, even if they are contextually novel.

This propensity for individuals with schizophrenia to be overly reliant on familiarity-based memory processes, manifesting as increased susceptibility to false recognition judgments, may extend to some extent to members of the general population exhibiting positive symptoms of schizotypy. Bhatt et al. (2010) observed a positive correlation between number of false positives in a memory task and delusionality within a healthy sample used as a control group for a patient study. Evans et al. (2019) found an association between delusionality and false pictorial memory (i.e., a high-confidence "old" response for "new" images) among college students—no parallel associations were found with negative symptoms or symptoms of disorganization. This genre of memory distortion has also been observed among individuals from the general population who report anomalous experiences such as alien abduction (Clancy et al., 2002) – in this case, symptoms of schizotypy were significant predictors of false recognition rate. Finally, as previously noted, Sahakyan and Kwapil (2019) found that positive, but not negative, schizotypy was associated with an increased rate of false alarms in recognition memory despite a stable hit rate—a finding that is indicative of elevated susceptibility to false recognition that cannot be explained by a general positive response bias. Together, these studies suggest that novelty detection deficits are dimensional in nature and may be most closely associated with positive schizotypy rather than being inherent to schizophrenia as a categorical disorder.

Overconfidence

An increased false recognition error rate may not be the only area of deficit that contributes to memory distortions among individuals experiencing positive symptoms of schizotypy. Several studies point to a dissociation between the *rate* of false memories reported and one's *confidence* in these false memories (Corlett et al., 2009; Dietrichkeit et al., 2020; Moritz et al., 2006). In these studies, delusion-prone individuals did not produce more false memories; they were, however, more confident in these incorrect responses. Such observations have led to a model of delusions that holds "liberal acceptance," or the propensity to require less information to come to a highly confident conclusion, as a core feature of delusions (Moritz et al., 2008). This theory has been corroborated by a number of studies reporting overconfidence in errors relative to correct responses among delusion-prone samples (for a review, see Balzan & Hodkinson, 2016). One could imagine how overconfidence might contribute to delusional beliefs such as paranoia by considering the fact that a high-confidence false memory (e.g., "I'm *sure* I put money in my wallet yesterday") is more likely to impact one's thoughts and behaviors than a low-confidence one (e.g., "I might have put money in my wallet yesterday, but it could have been last week"). In this way, overconfidence might contribute to a base of "evidence" (e.g., "Somebody must be stealing from me") that supports a persecutory belief (e.g., "People are out to get me").

Elevated confidence in erroneous judgments might also undermine the detection or incorporation of evidence that might otherwise disconfirm such a belief. Empirical support for a connection between metacognitive failure and a bias against disconfirmatory information can be found in Rollwage et al. (2018). In that study, individuals who displayed reduced metacognitive sensitivity in a perceptual decision-making task were also less responsive to post hoc information that disconfirmed their highly confident incorrect responses. These same individuals were more likely to endorse radical political beliefs. In similar fashion, an overly liberal metacognitive monitoring system may help to maintain other unconventional beliefs such as persecutory ideation by inflating confidence in interpretations that are in fact errors and making one resistant to corrective feedback regarding these errors (i.e., failing to recognize them as errors). Overconfidence is thus a metacognitive process that may contribute to (a) the *formation* of rigid persecutory beliefs by increasing the odds that a false memory is taken seriously by the paranoia-prone individual, as well as (b) the *maintenance* of such beliefs by decreasing the odds that the paranoia-prone individual revises their erroneous appraisals in the face of disconfirmatory information.

Based on the foregoing, we conjecture that overconfidence (as facilitated by poor metacognitive monitoring) represents a distinct, but interrelated, process that might combine with novelty detection deficits (as facilitated by an overreliance on familiarity-based memory) to produce not only an *elevated rate* of false recognition among paranoia-prone individuals, but also *heightened confidence* in these errors.

Present Study

Although past research provides tentative evidence for some level of dysfunction in both novelty detection and metacognitive monitoring among individuals experiencing positive symptoms of schizotypy, the distinct contributions of these processes to instances of false recognition and their relationship to paranoia remain to be elucidated. Furthermore, outside of the realm of DRM procedures and semantic association, little work has been devoted to understanding the circumstances under which false recognition might occur. To what extent are generally familiar (though semantically unrelated) stimuli sufficient to cause instances of false recognition? What might this tell us about the mechanisms that lead to memory distortion in daily life?

If paranoid individuals rely on a context-agnostic, familiarity-based memory style when making recognition judgments, one would expect memory performance to track with the extent to which context-specificity is necessary to make a correct rejection. In other words, one would expect novelty detection to be more impaired for lures which elicit a general, task-irrelevant sense of familiarity (e.g., items which are encountered with high frequency in everyday life) relative to stimuli that are accompanied by no such sense of familiarity (e.g., items to which one has had no prior exposure). In this study we manipulated familiarity through the use of stimuli that varied in their rate of extraexperimental exposure (i.e., real words vs. pseudowords). Further, to capture the magnitude of overall disruption to both memory and metacognitive performance, we calculated a hierarchical Bayesian estimate of meta-d' (Hmeta-d; Fleming, 2017) in addition to the more classic d'. First developed by Maniscalco and Lau (2012), meta-d' is a metric that has been used to index metacognition across a number of domains, including perception and memory (see Rouault et al., 2018 for a review). As the confidence-by-stimulus-by-response matrices that underlie meta-d' estimations involve confidence ratings weighted by the number of responses of a given type (i.e., hits, misses, correct rejections, false alarms; see Methods section for more detail), meta-d' scores are uniquely poised to index the extent to which the two memory distortions of interest (elevated false alarm rate and overconfidence) might synergize among paranoid individuals. To our knowledge, this is the first study to take this modeling approach to study metacognitive deficits associated with paranoia.

We tested the following hypotheses:

1. Paranoia will be associated with reduced metacognitive sensitivity (indexed by meta-d'); this effect will be accentuated for real-word relative to pseudoword trials.

2. Paranoia will be associated with an elevated false alarm rate; this effect will be accentuated for real-word relative to pseudoword trials.

3. Paranoia will be associated with elevated confidence on errors (i.e., false alarms and misses) on both real- and pseudoword trials.

Method

Data Collection

Data collection occurred in two waves. The hypotheses that developed based on analysis of wave 1 were preregistered (https://osf.io/hnj2w) prior to the collection of wave 2. For the sake of clarity and brevity, and in order to take advantage of the largest sample available, the main body of this article presents an omnibus analysis that combines data from wave 1 (not preregistered) and wave 2 (preregistered). Individual analyses of wave 1 and wave 2, and a more detailed account of areas of both convergence and divergence between them, can be found in Section S4 of online supplementary material.

Participants

392 participants were recruited to take an online survey via Amazon's Mechanical Turk (MTurk) platform (wave 1: $n = 117$, wave 2: $n = 275$). Only participants who were over the age of 18 and who were located in the United States were recruited. Following exclusions (described in detail below), wave 1 participants included 55 males and 35 females whose average age was 36.51 ($SD = 12.10$). Sixty-two (68.89%) participants reported having received a baccalaureate or postbaccalaureate degree. Wave 2 participants included 115 males and 112 females whose average age was 39.42 ($SD = 12.19$). One hundred thirty-five participants (59.47%) reported having received a baccalaureate or postbaccalaureate degree. See Section S1 of online supplementary material for more detailed demographic information.

Data Quality and Exclusion Criteria

A number of steps were taken to ensure data quality, in line with methods described in the wave 2 preregistration. First, surveys were only open to workers with a HIT approval rate above 95% and who had completed more than 100 HITS. Studies employing workers with approval rates above 95% have obtained results in online surveys that are comparable to parallel studies conducted in a traditional laboratory setting (Johnson & Borden, 2012). Additionally, to prevent multiple responses from a single MTurk worker, only one response was accepted per MTurk ID. Finally, at the end of the surveys, participants responded to a screener question querying random responding. Participants who responded

"Yes" to this question were excluded from further analysis. Additionally, participants were excluded from further analysis if (a) they did not finish the survey, (b) they responded to less than 50% of trials on the memory task, or (c) percent accuracy on the memory task was greater than 1 *SD* below chance-level responding (i.e., below 36.46% in wave 1; 36.08% in wave 2). This latter exclusion criterion was employed to avoid excluding participants who might have systematic memory distortions (i.e., the processes of interest in the present study) that negatively impact their performance, while still screening out participants who were entirely unengaged in the task. This criterion excluded 13 participants in total. Importantly, all patterns described in the Results section held qualitatively (and, in the case of the false alarm binary logistic regression, in terms of significance as well) when excluding all participants who performed below chance. In the second wave, participants were also asked to self-report their primary language. All participants reported their primary language to be English—this established a baseline level of exposure to the common English words used for the word familiarity manipulation (see Measures section for more information). Following exclusions, the final sample included 317 participants (wave 1: $n = 90$, wave 2: $n = 227$).

Measures

Questionnaires

In the online surveys, participants completed the Revised Green Paranoid Thoughts Scale (R-GPTS; Freeman et al., 2019). The R-GPTS consists of a 10-item persecution subscale and an 8-item reference subscale, both of which query thoughts and feelings one may have had about others in the past month. In both waves of data collection, paranoia was measured using the R-GPTS persecution subscale. In wave 2, referential thinking was measured using the R-GPTS reference subscale as an exploratory pilot to inform future research. As this article focuses on the persecution subscale, all additional information and analyses regarding the reference subscale can be found in online supplementary material.

The R-GPTS persecution subscale includes statements such as "I was convinced there was a conspiracy against me". Participants were instructed to indicate the extent to which they experienced these feelings on a scale of 0 (*not at all*) to 4 (*totally*). The sum of participants' responses on the persecution subscale was used to index paranoid ideation. In wave 1, missing data points were imputed using the mean of all other responses on the persecution subscale. No participant had more than a single missing item. In wave 2, there were no missing responses on this questionnaire.

In wave 1, a total of 36 participants, or 40.00% of the sample, scored above the threshold recommended by Freeman et al. (2019) to indicate moderate paranoia (11); in wave 2, it was 70 participants, or 30.84% of the sample. Although heightened rates of psychopathology are not unheard of on MTurk (Ophir et al., 2020), it is important to point out the large proportion of participants above threshold for moderate paranoia in the wave 1 sample. Despite being nonclinically ascertained, this sample may not be representative of a random sample from the general population and may be better understood as having oversampled individuals high on paranoia. The internal consistency of questionnaire measures was indexed using Omega total (McDonald, 1999). This metric is the result of a factor analysis of all items on a scale, followed by an oblique rotation and extraction of a general factor. Compared to Cronbach's alpha, Omega total has the advantage of accounting for strength of associations between items in addition to measurement error on an item-specific level. It can be interpreted using similar cut-offs as Cronbach's alpha, with a value above 0.9 reflecting excellent internal consistency. Descriptive statistics and Omega Total for questionnaires (wave 1, wave 2, total) can be found in Table 1.

Recognition Memory Task

In order to assess the effect of word familiarity on novelty detection, a recognition memory task was created, comprising real words—words in common use that come from the English language (e.g., "bike") – and pseudowords—words that resemble English words but that have no meaning (e.g., "rimp"). Mean accuracy across waves was 63.28% ($SD = 13.49$).

Stimulus Selection

Real words and pseudowords were selected using the English Lexicon Project website (https://elexicon.wustl.edu/; Balota et al., 2007). All words contained four letters and were matched across condition (real, pseudo) and group (target, lure) such that there were no significant differences in terms of a number of lexical characteristics. See Section S6 of online supplementary material for more information on stimuli selection as well as for a full list of stimuli used in this task.

Task Design

During Encoding, 40 items (20 real targets, 20 pseudo targets) were presented to participants in random order for 4 s each. Participants were instructed to remember as many items as possible and were informed that their memory of these items would later be tested. During Recognition, participants were presented with a total of 80 items (40 real, 40 pseudo) in random order. Half of these words had been presented during Encoding (i.e., targets); the other half had not (i.e., lures). For each item, participants were asked whether they thought it was "old" (i.e., presented during Encoding) or "new" (i.e., not presented during Encoding). Then, participants were asked to rate their confidence in their response on a scale of 1 (*very unsure*) to 4 (*very sure*). During Recognition, the survey auto-advanced once a response was logged or after 4 s (in wave 1) or 4.5 s (in wave 2) had elapsed. Comprehension questions before Recognition ensured that participants understood the distinction between "old" and "new"—the survey would not proceed until several practice trials were answered correctly.

Metacognitive Sensitivity

Recent metrics of metacognitive sensitivity have built upon d' – a signal detection theory (SDT) index of task performance. In the

Table 1
Descriptive Statistics and Omega Total of Questionnaires

Questionnaire	Wave 1	Wave 2	Total
R-GPTS—Persecution Subscale	11.26 (12.17)	7.99 (10.54)	8.92 (11.11)
	$\omega_t = .98$	$\omega_t = .97$	$\omega_t = .97$

Note. Descriptive statistics are reported as means (*SDs*).

context of a memory task, d' indexes "Type-I" memory performance, providing estimates of the distance between the means of one's internal "signal" and "noise" distributions. High d' values represent high discriminability, corresponding to a greater difference between these distributions and an increased likelihood of making a correct response. Meta-d', on the other hand, acts as a measure of "Type-II" performance, or the extent to which one's confidence ratings are predictive of one's actual success (Maniscalco & Lau, 2012, 2014).

One way to capture Type-II performance is "Hmeta-d" (Fleming, 2017) – a technique that uses hierarchical Bayesian methods to estimate meta-d' scores. In the Hmeta-d model, distributions of confidence ratings are generated conditional on whether one's response was correct or incorrect. A participant with higher metacognitive sensitivity is able to more accurately monitor their responses, providing higher confidence ratings on correct responses and lower confidence ratings on incorrect responses. In the model, this would result in weakly overlapping confidence distributions and a relatively high meta-d' score. Conversely, a participant with poor metacognitive sensitivity may endorse error trials with relatively high confidence or express reduced confidence on correct trials—this would result in highly overlapping confidence distributions and a relatively low meta-d' score. In the context of the present recognition memory task, which involved two categories of stimuli (i.e., target and lure), the inputs to the Hmeta-d model included one vector that denoted responses on trials where targets (i.e., old items) were presented—this vector indexed the number of hits (i.e., responses of "old") and misses (i.e., responses of "new") made by a participant, stratified by confidence ratings. The other vector did the same for lure trials (i.e., new items), indexing the number of correct rejections (i.e., responses of "new") and false alarms (i.e., responses of "old"), stratified by confidence ratings. Together, these vectors formed confidence-by-stimulus-by-response matrices that were the basis of HMeta-d model fit. In this way, HMeta-d efficiently incorporated several processes of interest to the present study, including false alarm rate and confidence allocation to errors versus correct responses.

Procedure

In both waves of data collection, participants completed Encoding before responding to the R-GPTS subscale(s). Participants then completed Recognition. Finally, participants responded to demographic questions.

Analyses

A preregistered data collection and analysis plan was filed after wave 1 had been collected and analyzed, but before wave 2 had been collected or analyzed. Note that the analyses presented below deviate from the preregistration in that they were carried out on both waves of data collapsed together rather than on individual waves. This was done to maximize power. This approach was justified in that there were no significant differences in overall performance between the first and second waves of data collection in terms of Type-I or Type-II SDT metrics (see Section S2 of online supplementary material). Individuals analyses of wave 1 and wave 2 can be found in Section S4 of online supplementary material. Finally, at the request of reviewers, this article reports continuous analyses of the paranoia variable rather than the categorical analyses that were planned in the preregistration.

Outliers were handled according to methods described in the preregistration. Outliers in nonskewed data were defined as points greater than 3 *SD*s from the sample mean; outliers in skewed data were defined using methods outlined by Hubert and Van der Veeken (2008), as implemented by R's RobustBase package (Todorov & Filzmoser, 2009). All identified outliers were winsorized (Fuller, 1991), preserving rank order. See Section S5 of online supplementary material for more information.

Trials were coded as a false alarm if participants responded "old" on a "new" trial; a correct rejection if "new" on a "new" trial. Trials were coded as a miss if participant responded "new" on an "old" trial; a hit if "old" on an "old" trial. Across confidence levels, these trials were used to calculate d' and beta (response bias) scores via the psycho package (Makowski, 2018). Confidence ratings were used to calculate meta-d' scores for each participant using methods described in Fleming (2017).

Our primary hypothesis was tested using a linear mixed-effects model of meta-d' scores using the lmer function of the lme4 package (Bates et al., 2015) in R. Models included an interaction term for paranoia by word familiarity condition (real vs. pseudo). For each model, age, sex, and education were included as covariates. The emmeans package (Lenth, 2020) was used to conduct all post hoc contrasts; Bonferroni correction was used to control for multiple comparisons. For completeness, the above models were also applied to d' and beta scores,

To probe the sources of group differences in meta-d' further, we performed a set of binary and ordinal logistic regressions on trial-by-trial responses and confidence ratings, respectively. Mixed-effects binary logistic regressions were carried out on the trial-by-trial counts (1s and 0s) of false alarms and misses from the recognition memory task. Note that correct rejections and hits are inverses of false alarms and misses, respectively; thus, no additional models were created for these response types. The mixed-effects binary logistic regressions were run in R via the glmer function of the lme4 package (Bates et al., 2015). Models included an interaction term for paranoia by word familiarity condition (real vs. pseudo). Mixed-effects ordinal logistic regressions were carried out on the trial-by-trial ratings of confidence for false alarms, correct rejections, misses, and hits. Confidence was specified as an ordinal variable, increasing in 1-unit increments from 1 to 4. The mixed-effects ordinal logistic regressions were run in R using the clmm function of the ordinal package (Christensen, 2019). Models included an interaction term for paranoia by word familiarity condition (real vs. pseudo).

Results

Zero-order correlations can be found in Section S7 of online supplementary material.

Linear Mixed-Effects Models

To test our primary hypothesis regarding an interaction between paranoia and word familiarity in the context of Type-II sensitivity, we created a linear mixed-effects model of meta-d' scores on real-word and pseudoword trials. Controlling for age, sex, and level of education, there were statistically significant main effects of paranoia, $t(471) = -2.64$, $p = .009$, $\eta_p^2 = 0.05$, 90% CI [0.02, 0.10], and of word familiarity, $t(315) = 5.15$, $p < .001$, $\eta_p^2 = 0.08$, 90% CI [0.04, 0.13], as well as a statistically significant interaction between paranoia and word familiarity, $t(315) = -2.20$, $p = .028$, $\eta_p^2 = 0.02$, 90% CI [0.00, 0.05]. As hypothesized, for each unit increase in paranoia, there was a greater decrease in meta-d' in the real-word condition versus the pseudoword condition, such that there was a smaller difference in Type-II performance between conditions among higher-paranoia individuals (see Figure 1A).

A similar analysis was carried out on d' scores in order to assess the relationship between paranoia and word familiarity in the context of Type-1 sensitivity. Controlling for age, sex, and level of education, there were statistically significant main effects of paranoia, $t(451) = -4.11$, $p < .001$, $\eta_p^2 = 0.09$, 90% CI [0.05, 0.15], and of word familiarity, $t(315) = 4.66$, $p < .001$, $\eta_p^2 = 0.06$, 90% CI [0.03, 0.11], as well as a statistically significant interaction between paranoia and word familiarity, $t(315) = -2.31$, $p = .022$, $\eta_p^2 = 0.02$, 90% CI [0.00, 0.05]. For each unit increase in paranoia, there was a greater decrease in d' in the real-word condition versus the pseudoword condition, such that there was a smaller difference in Type-I performance between conditions among higher-paranoia individuals (see Figure 1B).

Finally, a similar analysis was carried out on beta scores in order to assess the relationship between paranoia and word familiarity in the context of response bias. Controlling for age, sex, and level of education, there were no statistically significant effects (all p's > 0.21), suggesting that paranoia was not associated with a general bias to respond "old" versus "new" irrespective of stimulus identity.

Mixed-Effects Binary Logistic Regression

To test our second hypothesis—namely, whether paranoia was associated with an elevated false alarm rate, particularly among familiar stimuli—mixed-effects logistic regressions were carried out on the trial-by-trial counts of false alarms and misses. All models controlled for age, sex, and level of education. Pseudowords were used as the baseline category for the word familiarity term.

False Alarm

A mixed-effects logistic regression on false alarm trials revealed a statistically significant main effect of paranoia, $z = 6.53$, $p < .001$, and a statistically significant interaction between paranoia and word familiarity, $z = -3.96$, $p < .001$. The odds ratio was 1.04, 95% CI [1.03, 1.05] for the main effect, meaning that for each unit increase in paranoia, the odds of making a false alarm error increased by 1.04 times. The odds ratio was 0.98, 95% CI [0.98, 0.99] for the interaction term, meaning that—as hypothesized—for each unit increase in paranoia, there was a greater increase in proportion of false alarms in the real-word condition versus the pseudoword condition. This indicates that paranoia was associated with a greater likelihood of making a false alarm and that this effect was accentuated in the real-word condition (see Figure 2A).

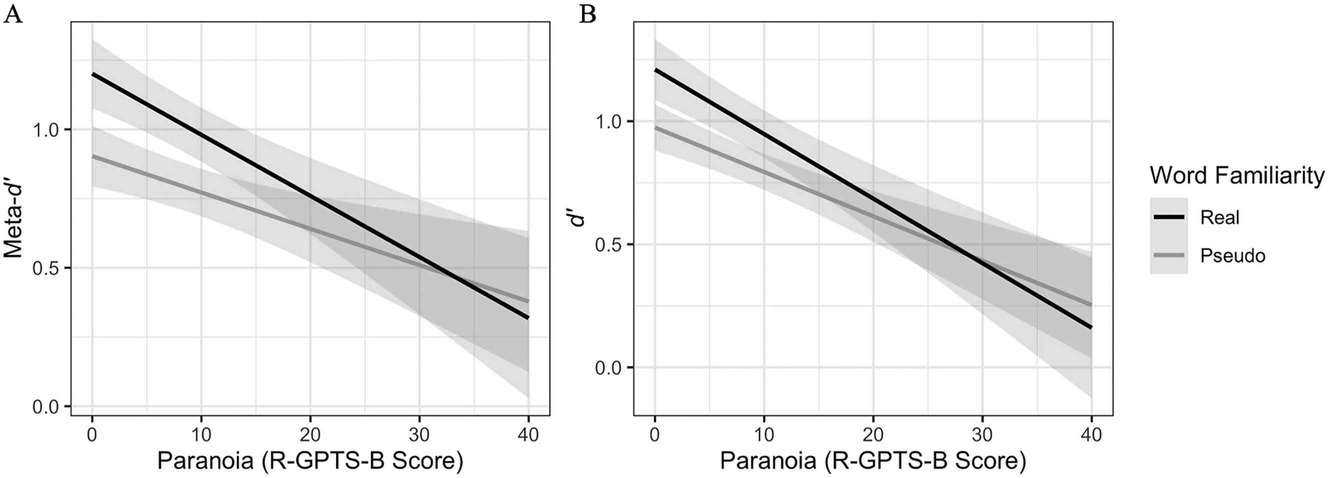
Miss

A mixed-effects logistic regression on miss trials revealed a statistically significant main effect of word familiarity, $z = 3.40$, $p < .001$. Neither a main effect of paranoia nor an interaction of paranoia by word familiarity was found ($p > .22$ for both effects). An odds ratio of 1.20, 95% CI [1.08, 1.34], indicated that the odds of responding with a miss in the pseudoword condition was 1.20 times that of the real-word condition, irrespective of paranoia status (see Figure 2B).

Mixed-Effects Ordinal Logistic Regression

To test our third hypothesis—namely, whether paranoia was associated with elevated confidence on error trials—mixed-effects ordinal logistic regressions were carried out on trial-by-trial ratings of confidence (1–4) on each response type (false alarm, correct rejection, miss, hit). All models controlled for age, sex, and level

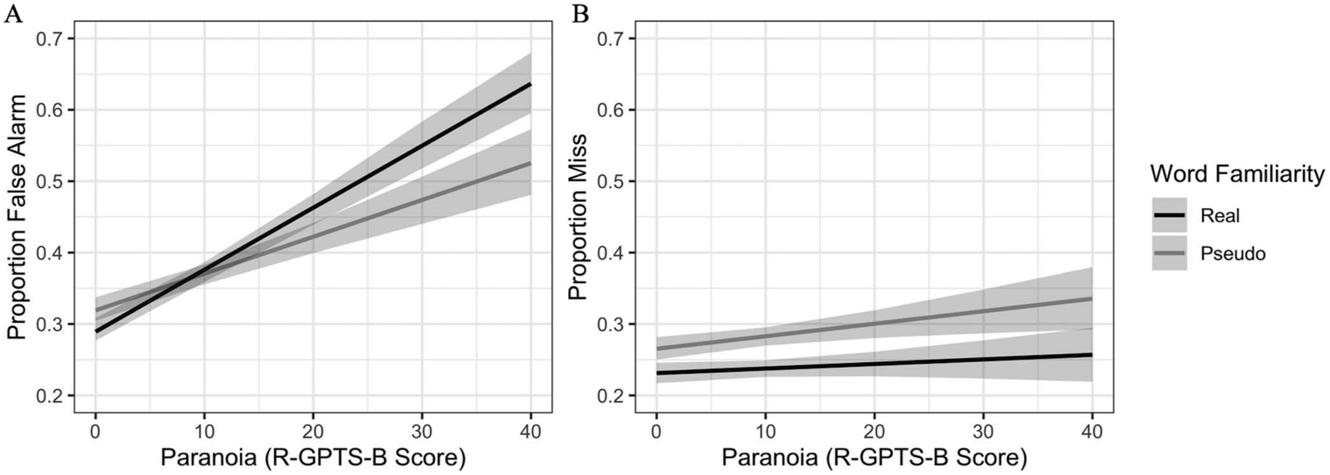
Figure 1
Meta-d′ (A) and d′ (B) as a Function of Paranoia and Word Familiarity Condition



Note. Shaded regions represent *SE*.

Figure 2

Proportion False Alarm (A) and Proportion Miss (B) as a Function of Paranoia and Word Familiarity Condition



Note. Shaded regions represent *SE*.

of education. Pseudowords were used as the baseline category for the word familiarity term.

False Alarm

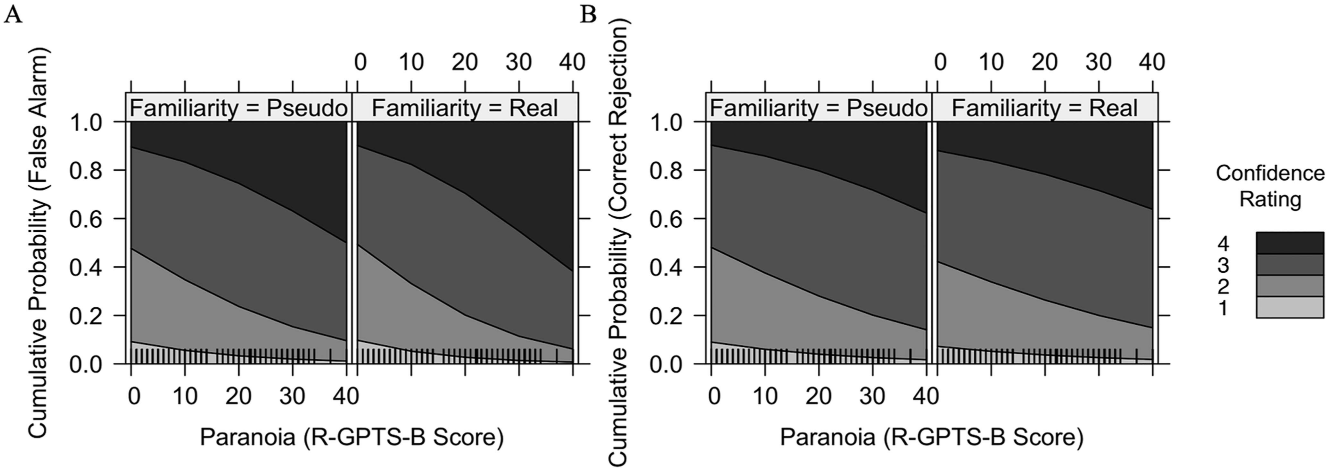
A mixed-effects logistic ordinal regression on confidence ratings for false alarm trials revealed a statistically significant main effect of paranoia, $z = 7.46$, $p < .001$, and a statistically significant interaction between paranoia and word familiarity, $z = -2.52$, $p = .012$. The odds ratio was 1.07, 95% CI [1.05, 1.09] for the main effect, meaning that for each unit increase in paranoia, the odds of endorsing a false alarm error with higher confidence increased by 1.07 times. The odds ratio was 0.99, 95% CI [0.98, 0.99] for the interaction term, meaning that for each unit increase in paranoia, there was a greater increase in the odds of making a higher-confidence false alarm in the real-word condition versus the pseudoword condition. This indicates that higher-paranoia individuals were more likely to endorse their false alarms with higher confidence and that this effect was accentuated in the real-word condition (see Figure 3A).

Correct Rejection

A mixed-effects logistic ordinal regression on confidence ratings for correct rejection trials revealed a statistically significant main effect of both paranoia, $z = 3.75$, $p < .001$, and word familiarity, $z = -4.20$ $p < .001$. No interaction was found ($p = .10$). The odds ratio was 1.04, 95% CI [1.02, 1.06] for the main effect of paranoia; 0.79, 95% CI [0.71, 0.88] for word familiarity. This indicates that for each unit increase in paranoia, the odds of

Figure 3

Cumulative Probability of Confidence Rating on False Alarms (A) and Correct Rejections (B) as a Function of Paranoia and Word Familiarity Condition



Note. Confidence ratings ranged from 1 (*very unsure*) to 4 (*very sure*).

responding with higher confidence on a correct rejection trial increased by 1.04 times. Additionally, the odds of responding with higher confidence on a pseudoword word trial was 0.79 times that of a real-word trial, holding constant all other variables (see Figure 3B).

Miss

A mixed-effects logistic ordinal regression on confidence ratings for miss trials revealed a statistically significant main effect of paranoia, $z = 4.76$, $p < .001$. No interaction was found ($p = .68$). An odds ratio of 1.05, 95% CI [1.03, 1.07], indicated that for each unit increase in paranoia, the odds of responding with higher confidence on a miss trial increased by 1.05 times (see Figure 4A).

Hit

A mixed-effects logistic ordinal regression on confidence ratings for hit trials revealed a statistically significant main effect of word familiarity, $z = -8.63$, $p < .001$. No interaction was found ($p = .15$). An odds ratio of 0.61, 95% CI [0.54, 0.68], indicated that the odds of responding with higher confidence on a hit trial in the pseudoword condition was 0.61 times that of the real-word condition, holding constant all other variables (see Figure 4B).

Discussion

Impaired Novelty Detection and Overconfidence

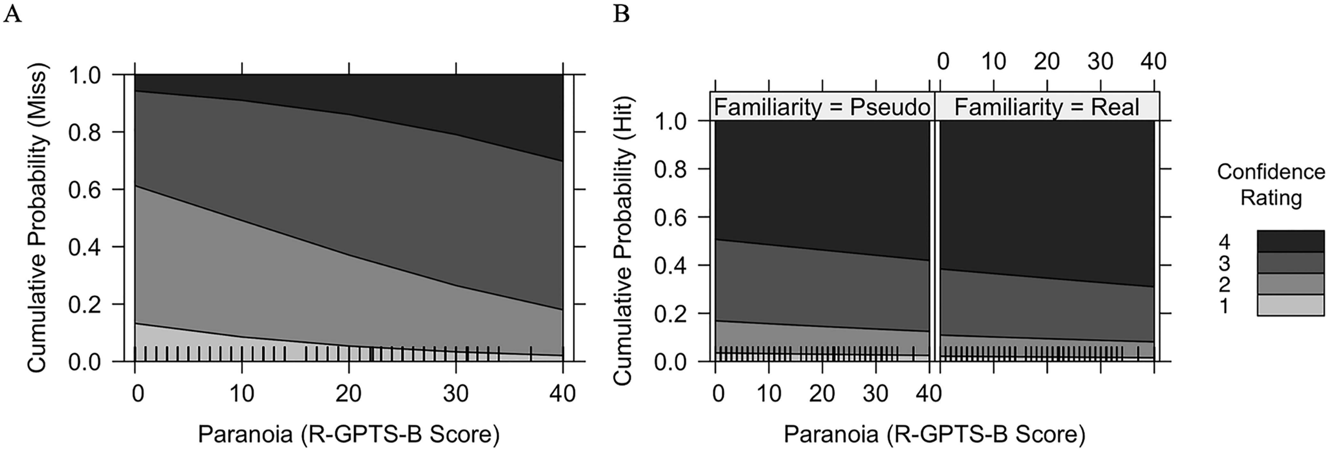
The present study provides strong evidence for an association between novelty detection deficits and paranoia in a general population sample, suggesting that persecutory ideation is associated with a heightened tendency to judge a new stimulus as having been previously encountered. The strength of this effect was striking, with higher-paranoia individuals approaching false alarm rates of 60% or more on certain conditions (e.g., real-word lures). Importantly, no parallel pattern was seen on miss trials. This level of specificity indicates that the generally poor memory performance of higher-paranoia individuals was largely driven by errors of commission (i.e., false alarms). Thus, the associations between paranoia and performance were unlikely to be driven solely by lack of engagement or generally impaired memory on the part of higher-paranoia individuals, in which case one would expect to see a nonspecifically elevated error rate (misses and false alarms alike). These observations suggest that novelty detection impairment likely represents a deficit whose association lies with paranoia itself, rather than with behavioral correlates of paranoia such as amotivation or generally impaired cognition.

In addition to showing an elevated rate of false recognition, individuals espousing persecutory beliefs endorsed their errors with high confidence. Whereas their lower-paranoia counterparts were more hesitant to express high confidence on error trials, the confidence ratings of higher-paranoia individuals revealed a pattern of decreased sensitivity to their mistakes, in which both miss and false alarm responses were assigned high confidence. Unexpectedly, paranoia was also associated with heightened confidence on correct rejection trials, suggesting that the association between paranoia and elevated confidence may not be as specific to errors as past research might suggest (e.g., Balzan & Hodkinson, 2016). This may be a corollary feature of the association between paranoia and elevated false alarm rate (and thus *decreased* correct rejection rate; see the results of the binary logistic regressions): it could be that the limited set of lure trials that were correctly rejected were very clearly recollected. In other words, in line with the observed deficits in novelty detection, items that were truly novel may have needed to be accompanied by an abundance of evidence in order to be correctly rejected, and thus may have been more likely to be endorsed with high confidence.

Our results suggest that these two effects combine to lead paranoid individuals to report a heightened number of instances of false recognition that are accompanied by strong conviction—a composite effect that was captured here in indices of both Type-I and Type-II performance. These findings are indicative of a dual disruption of both memory and metacognitive monitoring systems

Figure 4

Cumulative Probability of Confidence Rating on Misses as a Function of Paranoia (A) and Hits as a Function of Paranoia and Word Familiarity Condition (B)



Note. Confidence ratings ranged from 1 (*very unsure*) to 4 (*very sure*).

among individuals high in paranoia—not only do such individuals experience greater memory distortion, they also show a decreased ability to appropriately allocate confidence between correct and incorrect responses. Importantly, all effects held when controlling for other factors that might impact memory performance, including age and level of education, suggesting that the observed patterns are not an artifact of group differences on demographic measures. Between the frequency of and confidence in novelty detection errors, paranoid individuals might be differentially impacted by this type of memory bias in terms of their thoughts, feelings, or behaviors on a day-to-day basis. Further, these memory distortions may be characterized by a resistance to correction engendered by poor metacognitive monitoring (as suggested by Rollwage et al., 2018).

These findings go beyond prior work on the association between false belief and novelty detection by identifying and dissociating the specific underlying processes (i.e., impaired novelty detection plus overconfidence) that contribute to memory distortions associated with paranoia. Further, the presence of these deficits among the general population suggests some level of dimensionality in the relationship between memory bias and fixed false belief. This interpretation is consistent with the idea that certain symptoms of schizophrenia may not represent categorical departures from "normal" functioning; rather, they may occupy an extreme position on a spectrum of biases and heuristics that are also present—albeit to a lesser degree or within more specific contexts—in nonclinical populations.

Familiarity-Based Memory

The present study also helped to establish the contexts in which the observed deficits might be relatively pronounced among higher-paranoia individuals. Interactions between paranoia and word familiarity among d' and meta-d' metrics indicated that both Type-I and Type-II performance were relatively more disrupted in the presence of familiar stimuli (i.e., real words) among individuals higher in paranoia. A closer examination of the data reveals this dissociation to be largely driven by an *increase* in performance on the part of lower-paranoia individuals, who appeared to receive some benefit to their ability to remember stimuli and allocate confidence to their responses in the real-word condition. This dissociation may be explained by considering the underlying memory systems that contribute to recognition judgments. For an individual who is able to engage in conscious recollection, familiar words might offer an advantage insofar as they map more readily onto existing constructs or mnemonic devices than do pseudowords. However, for an individual who relies less on conscious recollection and more on a feeling of familiarity elicited by a given item, any benefit that might be gained from a real-word study list is likely to be undermined by the general sense of familiarity that accompanies *all* real-word stimuli, target and lure alike. Put differently, such an individual would be led to falsely recognize a lure as having been previously presented if the nonspecific, extraexperimental sense of familiarity with which it was accompanied was sufficient to surpass their "old-new" criterion. Thus, the fact that higher-paranoia individuals *did not* seem to benefit from the familiarity of the study list offers some insight regarding the processes that might give rise to false recognition—namely, an overreliance on context-agnostic, familiarity-based memory. Consistent with this interpretation, binary logistic regression models revealed that interactions in overall memory performance were differentially driven by errors of commission in the real-word condition. Thus, for higher-paranoia individuals, performance among real words was specifically undermined by *overattribution of prior exposure*—exactly as would be expected if one was using a contextually nonspecific sense of familiarity as a stand-in for conscious recollection. Finally, a similar interaction effect emerged in the analysis of confidence ratings, in which real-word false alarms were differentially endorsed with higher confidence among higher-paranoia individuals. This suggests that the extraexperimental familiarity of the real-word lures may have driven those higher in paranoia to not only falsely recognize these lures at relatively higher rates, but also to feel more confident in these erroneous responses. Together, these findings are suggestive of an association between paranoia and the use of gist-based feelings of familiarity to inform both memory and metamemory judgments.

These effects suggest that items which are *contextually*, but not *absolutely*, novel might be the most difficult to recognize as "new" for individuals high in paranoia. Anecdotally, this genre of stimulus may be more common to daily life: one might be more likely to encounter a relatively familiar stimulus in a novel context, in which it might take on a new meaning, than one is of encountering an entirely unfamiliar item (which migrates to a state of familiarity following exposure). The detection of contextual versus absolute novelty might engage unique encoding functions—Kafkas and Montaldi (2018) suggest that absolute novelty triggers acetylcholine-mediated hippocampal encoding while contextual novelty relies on dopaminergic/noradrenergic systems that engage a hippocampal-midbrain circuit. Future research should address whether there are distinct deficits in the detection of contextual versus absolute novelty among delusion-prone individuals—aberrancies in these systems may lead to a decreased tendency for hippocampally mediated systems to flip into an "encoding" mode, thereby reducing effective learning and blocking the integration of new information with preexisting knowledge or past experience.

Limitations

This study had several limitations. First, although all of the main effects of paranoia showed full independent replication across waves, the interactions between paranoia and word familiarity did not quite reach statistical significance in wave 2 alone. However, interaction effects were close to significance in the case of the binary logistic regression and d' models (p's = 0.056, 0.069, respectively; see Section S4 of online supplementary material) and qualitative patterns were consistent with the effects found in both wave 1 and the collapsed sample. This suggests that the magnitude of familiarity effects was likely overestimated in wave 1. However, the fact that familiarity effects held in the largest and most representative sample (the collapsed paranoia sample) lends credence to the notion that, in general, a nonspecific sense of familiarity may be sufficient to produce elevated rates of false recognition and overconfidence among higher-paranoia individuals.

This study also shares the limitations of most studies carried out on MTurk. Given that the wave 1 sample appeared to be enriched for paranoia, it is especially important to remember that the extent to which MTurk workers are representative of the general population remains unclear and that estimates of psychopathology depend entirely on self-report. We cannot rule out the possibility of comorbidities or formal diagnoses of psychotic illness—thus, the specificity of reported effects to paranoia should be the subject of future research. Additionally, the ecological validity of a cross-sectional memory task is necessarily limited—future work should address how novelty detection deficits might affect behavior in daily life.

Finally, in light of the low overall performance on the task (63.28% across waves), it is important to consider that the effects outlined in this article may be enhanced when recognition is difficult—in other words, if overreliance on familiarity cues is a compensatory strategy that is differentially employed by those high in paranoia, then the interaction effects presented herein may be particularly pronounced under difficult task conditions and may be less evident under decreased memory load.

Implications and Conclusion

This study provided strong support for an association between novelty detection deficits and paranoia in the general population, highlighted two processes that contribute to this phenomenon, and offered preliminary evidence that such deficits may be relatively pronounced among familiar stimuli. If such memory errors are causally linked to paranoia as a dimensional construct, through either the development or maintenance of this symptom, then interventions that succeed in improving novelty detection or decreasing mnemonic overconfidence may be clinically relevant for a wide array of rigidly held beliefs, from feelings of suspicion to conspiracy theories to full-blown persecutory ideation. Metacognitive training (MCT) – in which one is alerted to and taught to challenge one's biases—is one example of such an intervention that has shown promising early results (Kumar et al., 2015; Liu et al., 2018). The present study began to disentangle the processes whose synergy may cause novelty detection systems to fail—a more nuanced understanding of these systems and their interplay may offer important guidance to the development and optimization of intervention protocols such as MCT.

Understanding the association between novelty detection and paranoia might also provide insight into other phenomena that frequently co-occur with general delusionality, such as a cognitive bias against disconfirmatory evidence. When considering such a bias, it is important to understand how the "evidence" itself may be experienced on the part of the delusional individual. Characterizing associations between false belief and novelty detection deficits may be key to understanding how new information which might otherwise disconfirm a strongly held belief fails to be detected in the first place. Such an effect may be further accentuated by overconfidence, which might override corrective feedback regarding these false perceptions. More generally, to the extent to which novelty detection is necessary to engage in new learning, overattribution of familiarity may interfere with one's ability to flexibly form new and update preexisting associations about one's environment. Finally, to the extent that novelty and surprise are dissociable (Barto et al., 2013), a heightened false alarm rate may correspond to a subjective experience characterized by a greater proportion of stimuli in one's environment feeling *surprisingly* familiar. In this way, false recognition could participate in experiences of aberrant salience—unexpectedly familiar stimuli may become incorporated into delusional belief structures that assign meaning to perceived patterns among items that are unrelated to one another or are incidental to the current context (i.e., apophenia). However, these ideas are speculative and should be the subject of further research.

In a time of unprecedented and indiscriminate exposure to new (and often false) information, understanding how individuals process novel stimuli is of particular interest. Deepening our understanding of the connection between memory bias and incorrigible belief may lead to new insights regarding how, why, and in what contexts an individual might espouse persecutory beliefs, fall for fake news, or endorse elaborate hoaxes that have very tangible real-world impacts (e.g., climate change denial, COVID-19 conspiracy theories).

References

Achim, A. M., & Lepage, M. (2005). Episodic memory-related activation in schizophrenia: Meta-analysis. *The British Journal of Psychiatry*, *187*(6), 500–509. https://doi.org/10.1192/bjp.187.6.500

Aleman, A., Hijman, R., de Haan, E. H. F., & Kahn, R. S. (1999). Memory impairment in schizophrenia: A meta-analysis. *The American Journal of Psychiatry*, *156*(9), 1358–1366.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Balzan, R. P., & Hodkinson, K. (2016). Overconfidence in psychosis: The foundation of delusional conviction? *Cogent Psychology*. Advance online publication. https://doi.org/10.1080/23311908.2015.1135855

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology*, *4*, 907. https://doi.org/10.3389/fpsyg.2013.00907

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. Advance online publication. https://doi.org/10.18637/jss.v067.i01

Bhatt, R., Laws, K. R., & McKenna, P. J. (2010, July 30). False memory in schizophrenia patients with and without delusions. *Psychiatry Research*, *178*(2), 260–265. https://doi.org/10.1016/j.psychres.2009.02.006

Bronstein, M. V., & Cannon, T. D. (2017). Bias against disconfirmatory evidence in a large nonclinical sample: Associations with schizotypy and delusional beliefs. *Journal of Experimental Psychopathology*, *8*(3), 288–302. https://doi.org/10.5127/jep.057516

Butler, R. W., & Braff, D. L. (1991). Delusions: A review and integration. *Schizophrenia Bulletin*, *17*(4), 633–647. https://doi.org/10.1093/schbul/17.4.633

Christensen, R. (2019). *Ordinal-regression models for ordinal data*. Retrieved from https://cran.r-project.org/package=ordinal

Clancy, S. A., McNally, R. J., Schacter, D. L., Lenzenweger, M. F., & Pitman, R. K. (2002). Memory distortion in people reporting abduction by aliens. *Journal of Abnormal Psychology*, *111*(3), 455–461. https://doi.org/10.1037/0021-843X.111.3.455

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbins, T. W., Bullmore, E. T., Dickinson, A., & Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. *Brain: A Journal of Neurology*, *130*(9), 2387–2400. https://doi.org/10.1093/brain/awm173

Corlett, P. R., Simons, J. S., Pigott, J. S., Gardner, J. M., Murray, G. K., Krystal, J. H., & Fletcher, P. C. (2009). Illusions and delusions: Relating experimentally induced false memories to anomalous experiences and ideas. *Frontiers in Behavioral Neuroscience*, *3*, 53. https://doi.org/10.3389/neuro.08.053.2009

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dew, I. T., & Cabeza, R. (2013). A broader view of perirhinal function: From recognition memory to fluency-based decisions. *The Journal of Neuroscience*, *33*(36), 14466–14474. https://doi.org/10.1523/JNEUROSCI.1413-13.2013

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, *11*(9), 379–386. https://doi.org/10.1016/j.tics.2007.08.001

Dietrichkeit, M., Grzella, K., Nagel, M., & Moritz, S. (2020). Using virtual reality to explore differences in memory biases and cognitive insight in people with psychosis and healthy controls. *Psychiatry Research*, *285*, 112787. https://doi.org/10.1016/j.psychres.2020.112787

Dudukovic, N. M., & Knowlton, B. J. (2006, June). Remember-Know judgments and retrieval of contextual details. *Acta Psychologica*, *122*(2), 160–173. https://doi.org/10.1016/j.actpsy.2005.11.002

Evans, L. H., McCann, H. M., Isgar, J. G., & Gaston, A. (2019). High delusional ideation is associated with false pictorial memory. *Journal of Behavior Therapy and Experimental Psychiatry*, *62*, 97–102. https://doi.org/10.1016/j.jbtep.2018.09.005

Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1), nix007. https://doi.org/10.1093/nc/nix007

Freeman, D. (2016). Persecutory delusions: A cognitive perspective on understanding and treatment. *The Lancet Psychiatry*, *3*(7), 685–692. https://doi.org/10.1016/S2215-0366(16)00066-3

Freeman, D., Loe, B. S., Kingdon, D., Startup, H., Molodynski, A., Rosebrock, L., Brown, P., Sheaves, B., Waite, F., & Bird, J. C. (2019). The revised Green et al., Paranoid Thoughts Scale (R-GPTS): Psychometric properties, severity ranges, and clinical cut-offs. *Psychological Medicine*. Advance online publication. https://doi.org/10.1017/S0033291719003155

Freeman, D., McManus, S., Brugha, T., Meltzer, H., Jenkins, R., & Bebbington, P. (2011). Concomitants of paranoia in the general population. *Psychological Medicine*, *41*(5), 923–936. https://doi.org/10.1017/S0033291710001546

Fuller, W. A. (1991). Simple estimators for the mean of skewed populations. *Statistica Sinica*, *1*(1), 137–158.

Guo, J. Y., Ragland, J. D., & Carter, C. S. (2019). Memory and cognition in schizophrenia. *Molecular Psychiatry*, *24*(5), 633–642. https://doi.org/10.1038/s41380-018-0231-1

Hubert, M., & Van der Veeken, S. (2008). Outlier detection for skewed data. *Journal of Chemometrics*, *22*(3–4), 235–246. https://doi.org/10.1002/cem.1123

Jessen, F., Scheef, L., Germeshausen, L., Tawo, Y., Kockler, M., Kuhn, K.-U., Maier, W., Schild, H. H., & Heun, R. (2003). Reduced hippocampal activation during encoding and recognition of words in schizophrenia patients. *The American Journal of Psychiatry*, *160*(7), 1305–1312. https://doi.org/10.1176/appi.ajp.160.7.1305

Johnson, D. R., & Borden, L. A. (2012). Participants at Your Fingertips. *Teaching of Psychology*, *39*(4), 245–251. https://doi.org/10.1177/0098628312456615

Kafkas, A., & Montaldi, D. (2018). How do memory systems detect and respond to novelty? *Neuroscience Letters*, *680*, 60–68. https://doi.org/10.1016/j.neulet.2018.01.053

Kumar, D., Menon, M., Moritz, S., & Woodward, T. S. (2015). Using the back door: Metacognitive training for psychosis. *Psychosis*, *7*(2), 166–178. https://doi.org/10.1080/17522439.2014.913073

Lenth, R. (2020). *Estimate marginal means, aka least-squares means*. Retrieved from https://cran.r-project.org/web/packages/emmeans. https://doi.org/10.1080/00031305.1980.10483031

Lenzenweger, M. F. (2010). *Schizotypy and schizophrenia: The view from experimental psychopathology*. Guilford Press.

Libby, L. A., Yonelinas, A. P., Ranganath, C., & Ragland, J. D. (2013). Recollection and familiarity in schizophrenia: A quantitative review. *Biological Psychiatry*, *73*(10), 944–950. https://doi.org/10.1016/j.biopsych.2012.10.027

Liu, Y.-C., Tang, C.-C., Hung, T.-T., Tsai, P.-C., & Lin, M.-F. (2018). The efficacy of metacognitive training for delusions in patients with schizophrenia: A meta-analysis of randomized controlled trials informs evidence-based practice. *Worldviews on Evidence-Based Nursing*, *15*(2), 130–139. https://doi.org/10.1111/wvn.12282

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: Meta-*d*′, response-specific meta-*d*′, and the unequal variance SDT model. In S. Fleming & C. Frith (Eds.), *The Cognitive Neuroscience of Metacognition* (pp. 25–66). Springer. https://doi.org/10.1007/978-3-642-45190-4_3

McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum Publishers.

Moritz, S., Woodward, T. S., Jelinek, L., & Klinge, R. (2008). Memory and metamemory in schizophrenia: A liberal acceptance account of psychosis. *Psychological Medicine*, *38*(6), 825–832. https://doi.org/10.1017/S0033291707002553

Moritz, S., Woodward, T. S., & Rodriguez-Raecke, R. (2006). Patients with schizophrenia do not produce more false memories than controls but are more confident in them. *Psychological Medicine*, *36*(5), 659–667. https://doi.org/10.1017/S0033291706007252

Ophir, Y., Sisso, I., Asterhan, C. S. C., Tikochinski, R., & Reichart, T. (2020). The Turker blues: Hidden factors behind increased depression rates among Amazon's mechanical Turkers. *Clinical Psychological Science*, *8*(1), 65–83. https://doi.org/10.1177/2167702619865973

Provenzano, F. A., Guo, J., Wall, M. M., Feng, X., Sigmon, H. C., Brucato, G., First, M. B., Rothman, D. L., Girgis, R. R., Lieberman, J. A., & Small, S. A. (2020). Hippocampal pathology in clinical high-risk patients and the onset of schizophrenia. *Biological Psychiatry*, *87*(3), 234–242. https://doi.org/10.1016/j.biopsych.2019.09.022

Ragland, J. D., Ranganath, C., Harms, M. P., Barch, D. M., Gold, J. M., Layher, E., Lesh, T. A., MacDonald III, A. W., Niendam, T. A., Phillips, J., Silverstein, S. M., Yonelinas, A. P., & Carter, C. S. (2015). Functional and neuroanatomic specificity of episodic memory dysfunction in schizophrenia: A functional magnetic resonance imaging study of the relational and iItem-specific encoding task. *Journal of the American Medical Association Psychiatry*, *72*(9), 909–916. https://doi.org/10.1001/jamapsychiatry.2015.0276

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology*, *28*(24), 4014–4021. https://doi.org/10.1016/j.cub.2018.10.053

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: Insights from individual differences and neuroimaging. *Personality Neuroscience*, *1*, Article e17. https://doi.org/10.1017/pen.2018.16

Sahakyan, L., & Kwapil, T. R. (2019). Hits and false alarms in recognition memory show differential impairment in positive and negative schizotypy. *Journal of Abnormal Psychology*, *128*(6), 633–643. https://doi.org/10.1037/abn0000441

Schott, B. H., Voss, M., Wagner, B., Wustenberg, T., Duzel, E., & Behr, J. (2015). Fronto-limbic novelty processing in acute psychosis: Disrupted relationship with memory performance and potential implications for delusions. *Frontiers in Behavioral Neuroscience*, *9*, 144. https://doi.org/10.3389/fnbeh.2015.00144

Tamminga, C. A., Thomas, B. P., Chin, R., Mihalakos, P., Youens, K., Wagner, A. D., & Preston, A. R. (2012). Hippocampal novelty activations in schizophrenia: Disease and medication effects. *Schizophrenia Research*, *138*(2–3), 157–163. https://doi.org/10.1016/j.schres.2012.03.019

Thakral, P. P., Yu, S. S., & Rugg, M. D. (2015). The hippocampus is sensitive to the mismatch in novelty between items and their contexts. *Brain Research*, *1602*, 144–152. https://doi.org/10.1016/j.brainres.2015.01.033

Todorov, V., & Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*. Advance online publication. https://doi.org/10.18637/jss.v032.i03

Tulving, E., Markowitsch, H. J., Craik, F. I. M., Habib, R., & Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex*, *6*(1), 71–79. https://doi.org/10.1093/cercor/6.1.71

van Erp, T. G. M., Lesh, T. A., Knowlton, B. J., Bearden, C. E., Hardt, M., Karlsgodt, K. H., Shirinyan, D., Rao, V., Green, M. F., Subotnik, K. L., Nuechterlein, I., & Cannon, T. D. (2008). Remember and know judgments during recognition in chronic schizophrenia. *Schizophrenia Research*, *100*(1–3), 181–190. https://doi.org/10.1016/j.schres.2007.09.021

Weiss, A. P., Zalesak, M., DeWitt, I., Goff, D., Kunkel, L., & Heckers, S. (2004). Impaired hippocampal function during the detection of novel words in schizophrenia. *Biological Psychiatry*, *55*(7), 668–675. https://doi.org/10.1016/j.biopsych.2004.01.004

Woodward, T. S., Buchy, L., Moritz, S., & Liotti, M. (2007). A bias against disconfirmatory evidence is associated with delusion proneness in a nonclinical sample. *Schizophrenia Bulletin*, *33*(4), 1023–1028. https://doi.org/10.1093/schbul/sbm013

Woodward, T. S., Moritz, S., & Chen, E. Y. (2006). The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions: A study in an Asian sample with first episode schizophrenia spectrum disorders. *Schizophrenia Research*, *83*(2–3), 297–298. https://doi.org/10.1016/j.schres.2006.01.015

Received June 3, 2020
Revision received October 27, 2020
Accepted November 20, 2020 ■