Northumbria Research Link

Northumbria University
NEWCASTLE

Reality monitoring performance and the role of visual imagery in visual hallucinations.

Charlotte Aynsworth [a,b]

Nazik Nemat [a,b]

Daniel Collerton [b]

David Smailes [a1]

Robert Dudley [a,b]

[a] School of Psychology, Newcastle University, Newcastle upon Tyne, UK

[b] Northumberland, Tyne and Wear NHS Foundation Trust, UK

Correspondence to: Robert Dudley, Early Intervention in Psychosis services, Tranwell Unit, Queen Elizabeth Hospital, Gateshead, NE10 9RW

Tel: +44(0) 191 441 6580 email rob.dudley@ncl.ac.uk

[1] Present address: School of Psychology, Sunderland University, UK

Abstract

Background: Auditory Hallucinations may arise from people confusing their own inner speech with external spoken speech. People with visual hallucinations (VH) may similarly confuse vivid mental imagery with external events. This paper reports two experiments exploring confusion between internal and external visual material.

Method: Experiment 1 examined reality monitoring in people with psychosis; those with visual hallucinations (n=16) and those without (n=15). Experiment 2 used two non-clinical groups of people with high or low predisposition to VH (HVH, n=26, LVH, n=21). All participants completed the same reality monitoring task. Participants in Experiment 2 also completed measures of imagery.

Results: Psychosis patients with VH demonstrated biased reality monitoring, where they misremembered items that had been presented as words as having been presented as pictures. Patients without VH did not show this bias. In Experiment 2, the HVH group demonstrated the same bias in reality monitoring that psychosis patients with VH had shown. The LVH group did not show this bias. In addition, the HVH group reported more vivid imagery and particularly more negative imagery.

Conclusions: Both studies found that people with visual hallucinations or prone-ness to such experiences confused their inner visual experiences with external images. Vivid imagery was also related to proneness to VH. Hence, vivid imagery and reality monitoring confusion could be contributory factors to understanding VH.

Keywords:

- Visual Hallucinations; Reality Monitoring; Mental Imagery; Psychosis.

Introduction

Visual hallucinations (VH) are ill understood in comparison to auditory hallucinations (AH) particularly in people with psychosis (McCarthy-Jones et al., 2017). Cognitive models propose that AH arise owing to inner experiences (thoughts or inner speech) being confused with external experiences (someone else's actual speech; Bentall, 1990). This inner-outer confusion is thought to result from reality monitoring difficulties, which is the ability to recognise whether information is a true perception or imagined (Johnson, Hashtroudi, & Lindsay, 1993).

Psychosis patients with current hallucinations have difficulties with reality monitoring in comparison to psychosis patients without hallucinations (Brookwell, Bentall, & Varese, 2013) and there is consistent evidence that internally generated experiences (inner speech) are misattributed to an external source in clinical populations with AH (Jones, 2010). By extension it has been proposed that VH arise owing to internal mental images being misattributed as external perceptions (Brébion et al., 2012).

In the only study to date that directly tests reality monitoring in psychosis patients with VH, Brébion, Ohlsen, Pilowsky and David (2008) compared the performance of psychosis patients with VH against clinical (psychosis patients without VH) and non-clinical controls. Participants were presented with word/picture items (e.g., the word CAR, or a picture of a bicycle). After a short delay participants had to indicate whether items read from a list (including distractor items) were previously presented as a picture, a word, or not at all. This first stage established recognition accuracy. The second stage tested reality monitoring; as only the original target items were read out and participants identified whether the items had been presented as a word or a picture.

In the recognition phase people without VH demonstrated a picture superiority effect, and recognised pictures better than words. However, participants with VH showed the opposite pattern and recognised words better than pictures (a word superiority effect). It was proposed that VH patients were more likely to develop a vivid image when presented with a word and that this accounts for the absence of the usual picture superiority effect. In the reality monitoring stage, patients with VH differed from the other groups, by making more misattributions of words to pictures than people who did not report VH. Once again this implied VH were associated with heightened visual processes whereby words generated images readily and so were more easily confused with actual externally presented images.

This is an important study, that may reveal processes that lead to the experience of VH, that could feasibly be the target of therapeutic interventions (Smailes, Alderson-Day, Fernyhough, McCarthy-Jones, & Dodgson, 2015), but the findings must be interpreted cautiously. First, only a small number of participants with VH (n=8) took part, meaning the finding may be unreliable. Second, the control group of psychosis patients without VH (n=33) included people with and without other forms of hallucinations. Thus, differences in reality monitoring between the 'VH present' and 'VH absent' groups could be owing to reduced frequency of any hallucinations in the 'VH absent' group, rather than specifically relating to the presence of VH. Brébion et al.'s (2008) findings, therefore, require replication in a larger sample of psychosis patients, using a 'VH absent' group who report hallucinations in another modality.

In addition, the reality monitoring literature on AH suggests that processes involved in the development of clinical AH are also involved in non-clinical AH-like experiences (Badcock & Hugdahl, 2012; Laröi et al., 2004; but see Garrison et al., 2016, for two non-replications of this finding). However, the equivalent domain specific misattribution has not yet been demonstrated in relation to non-clinical VH.

Finally, Brébion et al. (2008) supposes that performance on the task and the apparent confusion as to the origin of material is owing to vivid mental imagery. Imagery is an important process in experiencing hallucinations in clinical (Aleman, Böcker, Hijman, De Haan, & Kahn, 2004) and non-clinical groups (Aleman, Nieuwenstein, Böcker, & De Haan, 2000). However, imagery was not measured by Brebion et al. (2008) and so it is unclear if this may account for the findings in relation to VH.

This present research consists of two studies investigating psychological processes leading to VH. This is important as VH are associated with high levels of distress, and disability (Mueser, Bellack &Brady, 1990) and there is scant mention of how to treat VH in psychosis with either medication or psychological therapy (Wilson, Collerton, Christodoulides & Dudley, 2015). Experiment one is a replication of Brébion et al.'s (2008) study, comparing reality monitoring performance of psychosis patients with VH compared to patients with AH, but without VH. Experiment two extends Brébion et al.'s work (2008) by investigating reality monitoring in non-clinical participants who are predisposed to VH, and examines the same hypotheses as experiment one which were based on Brébion et al's (2008) findings. First, it was predicted that people without VH would show the picture superiority effect, whereas those with VH will show the opposite and will recognise words more than pictures (a word superiority effect). Second, it was predicted that participants with VH would demonstrate a bias in reality monitoring where they misremember items that were presented as words as having been presented as pictures, more than participants without VH.

Method

Participants

Thirty three participants were recruited (20 males, 13 females) from Early Intervention in Psychosis and Psychosis Community Mental Health teams. Two people were excluded owing to difficulty understanding task requirements and establishing group membership. The VH group consisted of 16 people (age *M* = 25.75, *SD* = 6.35, 7 M, 9 F). The non-VH group consisted of 15 people (age *M* = 26.33, *SD* = 9.12, 12 M, 3 F). All of those in the VH group also experienced AH, whereas the non-VH group experienced AH only. The groups did not differ in age (VH: *M* = 25.75, *SD* = 6.35, non-VH: *M* = 26.13, *SD* = 8.45), $t(29) = -0.14$, $p = .89$ but did for gender X^2 (1, N =31) = 4.38, $p = 0.04$. All of the non VH group and 10 in the VH group were on antipsychotic medication. Of the six in the VH who were not on antipsychotic medication; two were on antidepressants; two were not on medication and two were unable to verify if they were taking medication. Diagnosis included first episode psychosis, paranoid schizophrenia, psychosis not otherwise specified, emotionally unstable personality disorder, and bipolar disorder with psychotic features. The inclusion criteria were; that the person was aged 18 years or more, reported hallucinations within the last six weeks, was in receipt of care, could give capacity to consent, and where they were prescribed antipsychotic medication, this was stable for at least 3 months. Exclusion criteria were; a history of substance abuse in the preceding 6 months, drug-induced hallucinations or psychosis owing to brain injury or organic disorders.

Sample size considerations

Brébion et al. (2008) reported large effect sizes for the difference between VH and non-VH groups on the reality monitoring task (Cohen's d = 0.99). Using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) the required sample size for a mixed model ANOVA with between-within subjects interaction ($p = .05$), and *f* value of 0.25, was 34 (17 per group) with power of 0.87.

Measures

Scale for the Assessment of Positive Symptoms. (SAPS; Andreasen, 1984)

This is a clinician-rated scale assessing positive symptoms of schizophrenia. It rates experiences from '0-None' to '5-Severe'. The scale demonstrates good psychometric properties with Cronbach's alpha of .66-.87 and inter-rater reliability of .70-1 (Sajatovic & Rairez, 2012). For this study's purpose, only the 7-item hallucinations subscale was administered.

North East Visual Hallucination Interview. (NEVHI; Mosimann et al., 2008).

The NEVHI is a 20-item semi-structured interview to assess for phenomenology of VH and their emotional, social and behavioural impact. Responses are scored on a 3-point likert scale ranging from 0 (little effect) to 2 (negative impact). It demonstrates good reliability (α = .71; κ =.83) and content validity (Mosimann et al., 2008).

Psychotic Symptom Rating Scales. (PSYRATS; Haddock, McCarron, Tarrier & Faragher, 1988).

The PSYRATS is a clinician administered semi-structured interview used to assess AH and delusions. Only the AH subscale (11 items) was administered. It covers a number of areas including frequency, duration, intensity of distress and disruption to life. Drake, Haddock, Tarrier, Bentall and Lewis (2007) found good reliability (α =.63-.76, κ = .99-1) and concurrent validity (r =.81) with the Positive and Negative Symptom Scale (Kay, Fiszbein, & Opler, 1987).

Materials

Object reality monitoring task (Brébion et al., 2008)

Encoding stage. Participants were presented with 16 words and 16 pictures of objects across 16 categories (Battig & Montague, 1969). In each category, participants were shown one word and one picture e.g. the word "FLUTE" and the picture "PIANO" (see Figure 1). They had to correctly identify a similar object from the same category (musical instruments), such as a trumpet, to indicate that they recognised the target items. Participants were randomly assigned to one of two conditions (A and B); which used opposite combinations of pictures and words (the word "FLUTE" and the picture "PIANO" or word "PIANO" and picture of a "FLUTE).

FIGURE 1 ABOUT HERE PLEASE

Recognition stage. After a delay, the experimenter read out a list of the 32 target words (those presented earlier) and 32 distractors (objects not shown during the encoding stage). Participants had to identify whether they were shown the item earlier as a picture, as a word or not at all. If people had stated a distractor as an example during the encoding phase, this was replaced with a different distractor.

Reality monitoring task. Participants were then read out a list of all 32 target items, without distractors. They had to identify whether they were shown these as a word or a picture.

Procedure

Participants were approached by their Care Co-ordinators and provided with an information sheet. If interested, the researcher contacted participants, who all provided informed consent prior to testing. Participants completed the encoding phase. After a 10 minute delay, participants undertook the recognition phase and then after a further 10 minutes the reality monitoring task. During the intervals, participants completed unrelated tasks. Finally, participants completed the NEVHI, SAPS and PSYRATS before being debriefed. Testing occurred either at participants' homes or NHS services.

Ethics

Both projects reported were subject to independent peer review and registered with the Trust Research and Development department and for this study a favourable NHS Ethical opinion was received.

Results

Data integrity checks

Data was screened for missing data. All experimental task data was present. All outliers were winsorized to meet normal assumptions according to the Shapiro-Wilks test. One participant from the VH group did not complete the NEVHI and one from the non-VH group did not complete the PSYRATS; presence of hallucinations was confirmed through verbal discussions. Owing to small numbers, this data was not replaced and they were excluded from group comparisons on these variables.

Participant characteristics

Participants who reported VH had significantly higher scores on the PSYRATS-AH scale than those without VH, $t(28)$ 2.72, $p = .011$ (VH, $M = 32.4$, $SD = 3.4$, non-VH, $M = 28.6$, $SD = 4.35$). Only the VH group completed the NEVHI ($M = 21.93$, $SD = 4.43$) as this measures VH experiences within the last month. Two participants from the non-VH group had experienced VH in the past, but not within the last 4 months. All of the VH group scored ≥2 on the SAPS hallucination scale indicating significant impact of VH on their lives. Of the 14 scored in the non-VH group, 11 scored ≥2 and three scored one indicating mild impact of AH.

Recognition

Summary data, shown in table 1, indicated that both groups performed well on the recognition task. False recognition of items was rare.

TABLE 1 ABOUT HERE PLEASE

A mixed model ANOVA explored differences in the types of stimuli presented (correct recognition words x correct recognition pictures) across groups (VH x non-VH). There was no main effect of stimuli on recognition, $F(1, 29) = 1.78$, $p = .19$ or of group on recognition, $F(1, 29) = .41$, $p = .53$. There was no significant interaction of group on the type of stimuli, $F(1, 29) = .74$, $p = .40$, indicating recognition of stimuli did not differ across groups and hypothesis one was not supported.

The analysis was repeated but with false recognition of new items as the within participant variable. There was no main effect of type of stimuli on false recognition, $F(1, 29) = .16$, $p = .69$ or of group on false recognition, $F(1, 29) = 2.01$, $p = .167$. There was no significant interaction of group on the type of false recognition, $F(1, 29) = .19$, $p = .67$, indicating false recognition of stimuli did not differ across groups.

Reality monitoring

Summary data are shown in table 2[2]. In order to test the hypothesis that people with VH have a bias in reality monitoring a mixed model compared type of misattribution (words misattributed to pictures x pictures misattributed to words) across groups (VH x non-VH) was undertaken. No significant main effect of type of misattribution, $F(1, 29) = 2.42$, $p = .13$ or group was found, $F(1, 29) = .05$, $p = .83$.

TABLE 2 ABOUT HERE PLEASE

[2] The performance of the groups on the correct remembering of words and pictures was not reported here as it was not a specified hypothesis. For completeness however, it can be reported that there were no group or task main effects and no interaction.

A significant interaction of group by type of misattribution was found, $F(1, 29) = 7.95$, $p = .009$ indicating there was difference between misattribution of words to pictures compared to pictures to words between groups. To examine this interaction a series of group and within subject t tests were used. There was no difference between VH and non-VH groups on misattribution of words to pictures, $t(24.7) = 1.31$, $p = .21$ (hypothesis two). Nor was there a significant difference found between groups on misattribution of pictures to words, $t(29) = -1.7$, $p = .08$. Paired sample t-tests found the VH group made significantly more misattributions of words to pictures than those of pictures to words, $t(15) = 3.31$, $p = .005$. There was no significant difference between misattributions to pictures or to words within the AH group, $t(14) = .84$, $p = .42$.

Discussion

This study explored the cognitive mechanisms that lead people to experience VH. With regards the first hypothesis, predicting a word superiority effect in the VH group there was no difference in recognition of words or pictures between groups. In the reality monitoring task, the interaction revealed a difference in performance between VH and non VH groups, and that the participants with VH did seemingly show a bias in reality monitoring, as they had a greater relative tendency to misremember items that had been presented as words as having been presented as pictures, which participants without VH did not show. There was not a group difference in that the VH group did not misattribute more words to pictures than those without VH so there is not direct support the hypothesis based on Brébion et al.'s (2008) findings. However, the results may be broadly consistent with the theory proposed by Brébion et al. (2008) and the literature in AH (Brookwell, Bentall, & Varese, 2013; Jones, 2010) that suggests VH may be caused by a reality monitoring problem whereby a person struggles to separate self-generated, internal images from external events.

An alternative explanation is that participants with VH may generate more vivid imagery so that when they are shown a word, it conjured up a vivid mental image of this, which on later recall made it harder for them to distinguish whether they were presented with a picture of a word. Thus, performance of participants with VH may not reflect a 'pure' reality monitoring problem, but may instead reflect 'normal' reality monitoring that has been influenced by unusually vivid imagery. This account could be supported by direct and indirect evidence for the role of imagery. Direct evidence would come from data showing that whilst undertaking this reality monitoring task participants with VH reported generating visual images of the object that is presented as a word, whereas people without VH do not. An indirect test would be to examine general levels of vivid mental imagery in people with and without VH.

The findings need to be considered in light of obvious limitations. The VH group scored higher on the PSYRATS-AH indicating that they had both visions and voices, and in effect had more hallucinatory experiences overall which may account for the performance, rather than the specific contribution of VH. Furthermore, other potential confounds were not controlled for such as medication, severity or length of psychosis or comorbid psychopathology which also may have impacted on the results.

To our knowledge, this is only the second study to explore reality monitoring in a sample of psychosis patients with VH. While the results are not a replication of Brébion et al.'s (2008), in that there was no difference on the recognition stage (presence of a word superiority effect) and differences on the RM task were only conceptually consistent with their findings, it may still suggest that inner-outer confusion could contribute towards people's experiences of VH. In experiment 2, we examined whether non-clinical participants who were prone to VH-like experiences (a) demonstrated reality monitoring problems and (b) reported experiencing unusually vivid mental imagery.

Experiment Two

Around 6% of the general population report VHs (Kessler et al., 2005). Recruiting people with such experiences allows the exploration of underlying mechanisms that could contribute to people experiencing clinical levels of distressing symptoms (Linscott & van Os, 2013). Studies have suggested that many, but not all, of the reality monitoring processes that play a role in the development of clinical AH also play a role in the development of non-clinical, AH-like experiences (see Badcock & Hugdahl, 2012, for a review). However, there have been no such investigations of non-clinical VH-like experiences. Hence, it is important to demonstrate if the inner-outer confusion processes reported in experiment 1 are also evident in people with VH-like experiences.

The hypothesised role for mental imagery in relation to reality monitoring needs further examination. As noted, a key limitation of previous work is the absence of direct or indirect measures of mental imagery. Imagery is a very plausible mechanism for increasing confusion between inner and outer experiences, as imagery is the "re-creation of perceptual experience" (Pearson, Deeprose, Wallace-Hadrill, Heyes & Holmes, 2013, p. 3). A high frequency of visual images, a greater vividness of visual images and greater emotional/psychological responses to the images has been reported in people with schizophrenia (Aleman et al., 2004; Libby & Eibach, 2011; Sack, Ven de Van, Etschenberg, Schatz, & Linden, 2005). Recently, Brébion et al., (2015) reported that people with VH had enhanced recognition of colour pictures, compared to people with AH, indicating that perception of visual images could be more vivid for people with VH (Oertel, Rotarska-Jagiela, Van de Ven, & Haenschel, 2009); thus increased vivid imagery could be specific to hallucinatory modality (Aleman et al., 2000).

Thus, visual imagery may be atypical in people with psychosis generally, and particularly those with VH. However, visual imagery varies across a number of dimensions including; frequency, valence, and vividness, each of which may be important in explaining why some people experience VH. For example, if the quality of imagery is experienced as very vivid by those predisposed to VH, then it may be that images are more readily confused with reality. Or it may only be when imagery is atypical in several different ways that VH are likely to occur.

In summary, this study is a replication and extension of the work of Brébion et al. (2008) in that it investigates reality monitoring in non-clinical participants who report VH-like experiences and examines the role of imagery. There are three main hypotheses. First, despite the lack of replication in study one, in keeping with the findings of Brebion et al. (2008) it was predicted that people with a higher predisposition to VH (HVH) would show a word superiority effect whereas those low in predisposition (LVH) will show a picture superiority effect. Second, for the reality monitoring stage it was predicted that HVH participants would misattribute more words as pictures than LVH as this is the crucial test of the misattribution theory. Third, in relation to imagery it was predicted that the HVH would demonstrate higher levels of spontaneous use of imagery, greater vividness of current imagery, and greater vividness of imagery for future events.

Method

This study consisted of two phases. A screening phase recruited university students to identify participants with high and low levels of VH. In the subsequent experimental task the two groups undertook a number of measures of imagery and a reality monitoring task.

Participants

Five hundred and twenty two students from two universities were recruited for phase one; 144 males and 371 females, aged 18-63yrs (M = 22.34, SD = 6.21). From this sample, participants were recruited based upon their responses to a measure of hallucination experiences (see measures section). The HVH consisted of 26 people (18-54yrs, M = 25.53, SD = 10.55, 5 M, 21 F) and the LVH consisted of 21 people (18-56yrs, M = 23.52, SD = 9.15, 4 M, 17 F). The groups differed in their reporting of VH experiences (HVH M = 9.50, SD = 1.30, LVH M= 5.00, SD= 0.00, ($t(45) = 17.60$, $p < 0.001$) but there was no difference in age ($t(45) = 0.69$, $p = 0.50$) or gender (χ^2 (1) = 0.00, $p = 0.99$)

Sample size considerations

Given the use of a non-clinical sample, a small to moderate effect size was expected (f=0.03). G*power estimated total group size of 55 was needed to achieve a high power (80%; Faul et al., 2007).

Measures

Launay Slade Hallucination Scale (Morrison, Wells & Nothard, 2000)

This is a 16-item self report measure of auditory and visual hallucination type experiences, using a 4-point Likert scale (1 = never, 4 = almost always). Higher scores indicating a higher predisposition to hallucinations. Based on a factor analysis that revealed two factor structure of auditory items, and visual items (Morrison et al. 2000) five items were used to assess predisposition to VH and to create the HVH and LVH groups. The items were 10) "I have seen a person's face in front of me when no one was there; 11) "When I look at things they appear strange to me"; 14) " I see shadows and shapes when there is nothing there"; 15) "When I look at things they look unreal to me" and 16) "When I look at myself in the mirror I look different".

On this VH subscale, scores range from 5 to 20, (for the sample in phase 1 n=522, $M = 6.93$, $SD = 1.88$). Hence, 1 SD above and below the mean was used as a cut-off for each grouping. A score of 5 was used for LVH and a score of 9 or more was used for HVH. However, owing to difficulties with recruitment this was reduced to 8, with no participants scoring below this. Good internal reliability for the entire sample on the VH subscale was demonstrated (α = .84).

Spontaneous Use of Imagery Scale (SUIS) (Reisberg, Pearson, & Kosslyn, 2003)

This 12-item measure rates frequency of imagery in day-to-day situations e.g. "If I saw a car accident, I would visualize what had happened when later trying to recall the details" on a Likert scale of 1-5 (1 = never appropriate, 5 = completely appropriate). Total scores range from 12 - 60, with higher scores indicating a higher use of general visual imagery. Nelis, Holmes, Griffith, and Raes' (2014) evaluation reported acceptable reliability and convergent validity. Good internal reliability for the experimental study was found (n= 47, α =.74).

Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973)

This 4-item measure asks participants to form images of a series of scenes (e.g., "Think of a relative or friend that you see frequently"), and then asks participants to rate the vividness of the details of the image (e.g., "The exact contour of face, head, shoulders, and body") on a Likert scale 1-5 (1 = perfectly clear; 5 = no image at all). Scores range from 16 to 80, with lower total scores indicating higher vividness of imagery. Good reliability was reported for this current study (n=47, α =.73).

Prospective Imagery Task (PIT, Stober, 2000; Holmes et al., 2011)

This is a 20-item measure of imagery of future positive (10 items; e.g. “People will admire you”) and negative (10 items; e.g. “Someone close to you will reject you”) scenarios and associated vividness. We used the version adapted by Holmes et al. (2011) with 3 Likert scales (vividness, likelihood, experiencing) to rate answers.

Ratings are given on a Vividness subscale (1 = no image; 5 = perfectly clear), a Likelihood subscale (1 = not likely to occur; 5 = very likely to occur), and an Experiencing subscale (1 = not at all; 5 = completely experiencing). For each subscale, scores are calculated for positive and negative items separately and range from 10 to 50. Higher scores indicate higher rates of vividness, stronger belief in the likelihood of the events happening and more intense experiencing of the emotions attached to the imagery. There was acceptable scale reliability for the positive (vividness subscale α =.73, likelihood subscale α = .73 experiencing subscale α =.75) and negative (vividness subscale α =.75 likelihood subscale α =.75, experiencing subscale α =.76) subscales in this study.

Reality Monitoring Task (Brébion et al., 2008)

This is the same task as used in Experiment 1.

Procedure

In the screening phase participants were given an information sheet and consent form and then completed the LSHS and SUIS. Participants who were eligible to take part in the experimental task were contacted. At testing they were provided with a second, study specific, information sheet and consent form. Participants completed the encoding stage of the reality monitoring task. They then completed the VVIQ and the PIT. After this delay period participants completed the recognition stage and then after a further interval during which they completed unrelated measures, the reality monitoring task.

Ethics

The study was approved by a University ethics panel.

Results

Data integrity checks

There was no missing data for the experimental task. Scores that were not normally distributed were adjusted by winsorizing the outliers.

Recognition

Data from Table 3 indicates that both groups made very few recognition errors and demonstrated a high level of accurate recognition.

TABLE 3 ABOUT HERE PLEASE

A mixed model ANOVA explored recognition performance (correct recognition words x correct recognition pictures) across groups (high X low). There was a significant main effect of stimuli, $F(1, 45) = 16.66$, $p < .001$, with pictures ($M = 13.66$, $SD = 1.75$) recognised more than words ($M = 11.92$, $SD = 2.79$). There was no main effect of group, $F(1, 45) = 0.04$, $p = .85$, or significant interaction, $F(1, 45) = 0.38$, $p = .54$, indicating both groups showed the picture superiority effect, once again not supporting hypothesis one.

The analysis was repeated but with false recognition of new items as the within participant variable. There was a significant effect of stimuli on false recognition scores, $F(1,45) = 17.89$, $p < .001$. Distractor items were more likely to be falsely recognised as words, rather than pictures (words $M = 1.41$, *std error* 0.29; pictures $M = 0.51$, *std error* 0.11). There was no effect of group $F(1,45) = 0.62$, $p = .44$, nor significant interaction of group on type of false recognition, $F(1,45) = 0.91$, $p = .35$

Reality Monitoring

TABLE FOUR ABOUT HERE PLEASE

As can be seen in Table 4 the participants were accurate in their identifications[3]. In order to test the second hypothesis a mixed model comparing type of misattribution (words misattributed for pictures X pictures misattributed for words) across groups (high X low) w was conducted. There was a significant effect of type of misattribution, $F(1,45) = 7.37$, $p = .01$. More words were mistaken for pictures, than pictures for words (respectively $M = 2.41$, $SD = 2.03$, $M = 1.59$, $SD = 1.34$). There was no effect of group $F(1,45) = 0.00$, $p = 1$ but there was a significant interaction of group on type of misattribution, $F(1,45) = 4.36$, $p = .04$, indicating that there were differences between HVH and LVH on types of mistakes.

Between subjects t-tests revealed that there were no significant differences between groups for words misattributed for pictures, $t(43) = 1.06$, $p = .29$, (hypothesis two) or pictures misattributed for words, $t(43) = -1.61$, $p = .11$. A paired-sample t-test for the low predisposition group, revealed no significant differences between words misattributed for pictures and pictures misattributed for words, $t(20) = 0.48$, $p = .64$. For the high predisposition group, there was a significant difference between words misattributed for pictures and pictures misattributed for words; $t(25) = 3.31$, $p = .003$, which is consistent with findings of study one.

Performance on Imagery measures

Performance on the imagery measures by the two groups is reported in Table 5.

TABLE FIVE ABOUT HERE PLEASE

[3] As with study one the analysis is not reported, as it was not a key hypothesis, however, for completeness no group main effects or interaction were found.

Between-groups t-tests revealed differences on the SUIS, t (45) = 15.78, $p < .001$, (Cohen's d= 0.60), but the not VVIQ, $t(45) = 1.66$, $p = 0.10$, d= 0.49. Analysis of participants' scores on the PIT was undertaken using a mixed 3 (dimension: vividness, likelihood, experiencing) × 2 (valence: positive, negative) × 2 (group: LPVH, HPVH) ANOVA. There was an effect of valence, $F(1, 45) = 52.79$, $p < .001$, with positive imagery rated higher (M= 35.31, SD = 5.06) than negative (M= 27.95, SD = 5.71). There was no group main effect, $F(1,45) = 2.38$, $p = .14$, or group by dimension interaction, $F(2,90) = 0.10$, $p = .91$, or group by valence by dimension interaction, $F(2, 90) = 0.07$, $p = .94$). However, there was a group by valence interaction, $F(1,45) = 4.27$, $p = 0.04$.

To explore this interaction between-subjects t-tests were undertaken on the total positive or negative imagery score (combined means for the three subscales). The groups did not differ on positive imagery (HVH M=105.54 SD= 14.53; LVH M= 106.33 SD= 15.77), $t(45) = 0.18$, $p = .86$, d = 0.05, but did differ on negative imagery (HVH M=89.73 SD= 15.50; LVH M= 77.95 SD= 18.73), $t(45) = 2.36$, $p = .023$, d = 0.66, with the HVH group reporting greater levels of negative imagery than the LVH group.

Discussion

This study investigated the role of reality monitoring and imagery in VH. No difference in recognition of words was evident between groups, once again providing no support for the first hypothesis. For hypothesis two, no group differences were found between the misattribution of words to pictures. However, as with our first study, the significant interaction indicated that people with HVH confused words for pictures more than they confused pictures for words. The low VH group did not show this pattern of errors. With regards to the third hypothesis, non-clinical participants who have higher levels of VH experiences reported using visual imagery more frequently in their everyday life, and in relation to negative situations, but not in terms of other aspects of imagery. Thus, the greater use of visual imagery may lead to the greater misattributions of internal images to external events in VH. However, these findings need to be replicated, as multiple comparisons were performed here, and it is possible that the group differences we have reported are Type I errors.

More broadly, various limitations need consideration. Owing to the low rate of VH type experiences in the non-clinical population the criterion for group membership meant that there was only modest difference between the groups. The measure used to identify the groups is widely-used but the LSHS items are very broad descriptions of unusual visual experiences, and focus on capturing the frequency rather than the intensity. The scale does not establish when or where the experiences happened (such as when waking up or falling asleep or whilst taking drugs) and so there are limitations when used to classify people in terms of their predisposition to VH. Furthermore, not all confounding factors were controlled for such as mood, or anxiety that could have impacted on group performance.

General discussion

Both studies presented here explored the confusion between internal experiences and external events by examining performance of people with VHs in clinical and non-clinical populations on a reality monitoring task. The study aimed to improve on past research by increasing the number of participants with VH, using a better matched control group and extending the work to non clinical participants. In light of these improvements, neither study provided support for the first hypothesis, that people with VH experiences will show a word rather than picture superiority effect at the recognition task. Therefore, we conclude that this was most likely an unreliable finding owing to the small sample size in the original Brébion et al (2008) study. In neither study did we replicate the between-group differences of greater misattribution of words to pictures by people with VH as in Brébion et al. (2008). Rather, both experiments reported an interaction in which clinical and non-clinical participants with VH experiences were more likely to misattribute words to pictures, than pictures to words. The non-VH groups did not demonstrate this, which suggests a unique pattern of misattribution in people with VH. These findings can, therefore, be considered consistent with models that propose atypical reality monitoring plays a role in the development of hallucinations (Bentall, 1990).

Reality monitoring differences are thought to occur because people with VH generate mental imagery in response to the presentation of a word, which leads to greater confusion when identifying the materials origin. We did not directly test whether the participants formed these images in relation to each presented word and we have only indirect evidence for the role of imagery as assessed by self-report measures. From Experiment Two it seems that visual mental imagery could be a contributor to this process but this requires future exploration in clinical samples, given the limitations noted earlier.

A general limitation of the work is the use of the Brébion et al. (2008) task. Whilst it has been used previously to examine RM performance (Brébion, Amador, David, Malaspina, Sharif & Gorman 2000) a particular issue is that the recognition stage also includes a source identification task as people not only say if the presented material is old or new but also if it is old, whether it was presented as a word or picture. Hence, it is not a pure recognition test, and means that material is presented twice to the person before the actual RM task is undertaken. Future studies may wish to only assess if the material is new or is previously presented to help reduce this potential confound. Another limitation to both studies is that neither adequately controlled for the possible contribution of mood or anxiety. Previous research (Brebion et al., 1997) with psychosis patients has indicated that the association between low mood and performance on a source memory task ($r = .17$) is weaker than the association between the association between positive symptoms and performance on a source memory task (r = .44). While this task is not identical to the task employed here, it supports the idea that the associations reported here would have persisted even after controlling for mood.

There are potential clinical implications of these findings. Currently, there is limited investigation of psychological treatments for VH, with only small case series exploring Cognitive Behavioural Therapy targeting VH (Wilson, Collerton, Christodoulides & Dudley, 2015; Thomson, Wilson, Collerton, Freeston & Dudley, submitted). Formulating VH in light of inner-outer confusion could develop more meaningful understandings of people's experiences, as applied in AH (Smailes et al., 2015) .

References

Andreasen, N. C. (1984). Scale for the Assessment of Positive Symptoms (SAPS). Iowa City, IA: University of Iowa.

Badcock, J. C., & Hugdahl, K. (2012). Cognitive mechanisms of auditory verbal hallucinations in psychotic and non-psychotic groups. *Neuroscience and Biobehavioral Reviews*. 36, 431–438. doi: 10.1016/j.neubiorev.2011.07.010

Battig, W.F., & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology (Monograph)* 80, 1-46. doi: 10.1037/h0027577

Bentall, R.P. (1990). The Illusion of Reality: A Review and Integration of Psychological Research on Hallucinations. *Psychological Bulletin*, 107, 82-95. PMCID: PMC2404293

Brébion, G., Ohlsen, RI., Pilowsky, LS., & David, A. S. (2008) Visual Hallucinations in schizophrenia: confusion between imagination and perception. *Neuropsychology, 22,* 383-389. doi: 10.1037/0894-4105.22.3.383

Brébion, G., Amador, X., David, A., Malaspina, D., Sharif, Z., & Gorman, J. M. (2000). Positive symptomatology and source monitoring failure in schizophrenia: An analysis of symptom specific effects. *Psychiatry Research, 95,* 119–131.

Brébion, G., Stephan-Otto, C., Usall, J., Huerta-Ramos, E., Perez del Olmo, M., Cuevas-Esteban, J., Haro, J.M., & Ochoa, S. (2015). Association of Auditory-Verbal and Visual Hallucinations with impaired and improved recognition of colored pictures. *Neuropsychology,* Advance online publication. doi: 10.1037/neu0000169

Brébion, G., Ohlsen, R., Bressan, A., & David. A.S. (2012). Source memory errors in schizophrenia, hallucinations and negative symptoms: a synthesis of research findings. *Psychological Medicine, 42,* 2543–2554. doi: 10.1017/S003329171200075X

Brookwell, M.L, Bentall, R.P., & Varese F.V. (2013) Externalizing biases and hallucinations in source monitoring, self-monitoring and signal detection studies: a meta analytical review. *Psychological Medicine, 43,* 2465-2475. doi: 10.1017/S0033291712002760

Drake, R., Haddock, G., Tarrier, N., Bentall, R., & Lewis, S. (2007). The Psychotic Symptom Rating Scales (PSYRATS): Their usefulness and properties in first episode psychosis. *Schizophrenia Research, 89*(1-3), 119-122. doi:10.1016/j.schres.2006.04.024

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

Garrison, J. R., Moseley, P. Alderson-Day, B., Smailes, D. Fernyhough, C. & Simons, J. S. (2016). Testing continuum models of psychosis: No reduction in source monitoring ability in healthy individuals prone to auditory hallucinations. *Cortex.* doi: 10.1016/j.cortex.2016.11.011

Haddock, G., McCarron, J., Tarrier, N., & Faragher, E. B. (1999). Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychological Medicine, 29*(4), 879-889. doi:10.1017/s0033291799008661

Holmes, E., Deeprose, C., Fairburn, C., Wallace-Hadrill, Bonsall, M.B., Geddes, J.R., Goodwin, G.M (2011) Mood stablitiy versus mood instability in bipolar disorder: A possible role for emotional imagery, *Behaviour Research and Therapy,* 1-6, DOI: 10/1016/J.BRAT.2011.06.008.

Jones, S. R. (2010). Do We Need Multiple Models of Auditory Verbal Hallucinations?

Examining the Phenomenological Fit of Cognitive and Neurological Models. *Schizophrenia Bulletin, 36,* 566-575. doi: 10.1093/schbul/sbn129

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin,* 114, 3-28.

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin, 13*(2), 261-276. doi:10.1093/schbul/13.2.261

Kessler, R. C., Birnbaum, H., Demler, O., Falloon, I. R. H., Gagnon, E., Guyer, M., … Wu, E. Q. (2005). The prevalence and correlates of non-affective psychosis in the National Comorbidity Survey Replication (NCS-R). *Biological Psychiatry*, *58*(8), 668–676. doi: 10.1016/j.biopsych.2005.04.034

Kosslyn, S.M., Ganis, G & Thompson, W. (2001). Neural Foundations of Imagery. *Nature Reviews: Neuroscience, 2,* 635-642. doi:10.1038/35090055

Laroi, F., Van der Linden, P., & Marczewski, P. (2004). The effects of emotional salience, cognitive effort and meta-cognitive beliefs on a reality monitoring task in hallucination-prone subjects. *British Journal of Clinical Psychology, 43,* 221-233. doi: 10.1348/0144665031752970

Libby, L.K., & Eibach, R.P. (2011). Visual Perspective in Mental Imagery: A Representational Tool that Functions in Judgment, Emotion, and Self-Insight. *Advances in Experimental Social Psychology*, *44*, 186-231. doi: 10.1016/B978-0-12-385522-0.00004-4

Linscott, R.J., & van Os, J. (2013). An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. Psychological Medicine 43, 1133–1149. doi: 10.1017/S0033291712001626

Marks, D.F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17-24. doi: 10.1111/j.2044-8295.1973.tb01322.x

McCarthy-Jones, S., Smailes, D., Corvin, A., Gill, M., Morris, D., Dinan, T., Murphy, K., O'Neill, F., Waddington, J., Australian Schizophrenia Research Bank, Donohoe, G., & Dudley, R. (2017) Occurrence and co-occurrence of hallucinations by modality in schizophrenia-spectrum disorders, Psychiatry Research, http://dx.doi.org/10.1016/j.psychres.2017.01.102

Mueser, K. T., Bellack, A. S., & Brady, E. U. (1990). Hallucinations in schizophrenia. *Acta Psychiatrica Scandinavica*, *82,* 26-29. doi:10.1111/j.1600-0447.1990.tb01350.x

Morrison, A.P., Wells, A., & Nothard, S. (2000). Cognitive factors in predisposition to auditory and visual hallucinations. *British Journal of Clinical Psychology*, 39, 67-78. doi: 10.1348/014466500163112

Mosimann, U. P., Collerton, D., Dudley, R., Meyer, T. D., Graham, G., Dean, J. L., . . . McKeith, I. G. (2008). A semi-structured interview to assess visual hallucinations in older people. *International Journal of Geriatric Psychiatry, 23*(7), 712-718. doi:10.1002/gps.1965

Nelis, S., Holmes, E. A., Griffith, J.W., & Raes, F. (2014). Mental imagery during daily life: Psychometric evaluation of the Spontaneous Use of Imagery Scale (SUIS*). Psychologica Belgica, 54(1),* 19-32. doi: 10.5334/pb.ag

Oertel, V., Rotarska-Jagiela, A., van de Ven, V., Haenschel, C., Grube, M., Stangier, U., . . . Linden, D. E. J. (2009). Mental imagery vividness as a trait marker across the schizophrenia spectrum. *Psychiatry Research, 167*(1-2), 1-11. doi:10.1016/j.psychres.2007.12.008

Pearson, D.G., Deeprose, C., Wallace-Hadrill, S.M.A., Heyes, S.B., & Holmes, E.A. (2013). Assessing mental imagery in clinical psychology: A review of imagery measures and a guiding framework. *Clinical Psychology Review, 33,* 1-23. doi: /10.1016/j.cpr.2012.09.001

Reisberg, D., Pearson, D.G., & Kosslyn, S. M. (2003). Intuitions and introspections about imagery: the role of imagery experience in shaping an investigator's theoretical views. *Applied Cognitive Psychology, 17,* 147–160. doi: 10.1002/acp.858

Sack, A., van de Ven, V., Etschenberg, S., Schatz, D., & Linden, D. (2005). Enhanced vividness of mental imagery as a trait marker of schizophrenia. *Schizophrenia Bulletin, 31*(1)*,* 1-8. doi: 10.1093/schbul/sbi011

Sajatovic, M., & Ramirez, L. F. (2012). *Rating Scales in Mental Health.* Michigan: JHU Press. doi: 1421406667, 9781421406664

Smailes, D., Alderson-Day, B., Fernyhough, C., McCarthy-Jones, S., & Dodgson, G. (2015). Tailoring Cognitive Behavioral Therapy to Subtypes of Voice-Hearing. *Frontiers in Psychology: Psychopathology, 6*, 1933. doi:10.3389/fpsyg.2015.01933

Stober, J. (2000). Prospective cognitions in anxiety and depression: a replication and methodological extension. *Cognition and Emotion,* 14 (5) 725-729. doi: 10.1080/02699930050117693

Thomson, C., Wilson, R., Collerton, D., Freeston, M., & Dudley. R. (2017). *Cognitive behavioural therapy for VHs: An investigation using a single case experimental design.* Manuscript submitted for publication.

Wilson, R., Collerton, D., Freeston, M., Christodoulides, T., & Dudley, R. (2015). Is seeing believing? The process of change during Cognitive Behavioural Therapy for distressing VHs. *Clinical Psychology and Psychotherapy.* doi:10.1002/cpp.1962

Table 1: Object recognition stage scores for both groups

	VH (*n* = 16)		Non VH (*n* = 15)	
	M	*SD*	*M*	*SD*
Correct words	12.94	1.52	12.13	2.97
Correct pictures	13.13	2.22	13	2.3
False recognition words	1	1.37	1.33	1.9
False recognition pictures	1.38	1.96	.67	.72

Note. Correct Words: word items correctly recognised out of 16 presented
Correct Pictures: picture items correctly recognised out of 16 presented
False recognition words: distractor items identified as words
False recognition pictures: distractor items identified as pictures

Table 2: Object reality monitoring task scores for both groups

	VH (*n* = 16)		Non VH (*n* = 15)	
	M	*SD*	*M*	*SD*
Correct words	11.00	2.82	12.27	2.15
Correct pictures	12.56	2.68	11.33	3.04
Words misattributed to pictures	5.31	3.42	4.00	2.04
Pictures misattributed to words	3.00	2.07	4.67	3.04

Table 3: Object recognition stage scores for both groups

	High Group *n* = 26		Low Group *n* = 21	
	M	SD	M	SD
Correct Words	11.84	3.17	12.00	2.42
Correct Pictures	13.85	1.93	13.48	1.57
False recognition words	1.19	1.29	1.62	1.83
False recognition pictures	0.50	0.76	0.52	0.68

Note. Correct Words: word items correctly recognised out of 16 presented
Correct Pictures: picture items correctly recognised out of 16 presented
False recognition words: distractor items identified as words
False recognition pictures: distractor items identified as pictures

Table 4: Object reality monitoring task scores for both groups

	High group (n = 26)		Low group (n = 21)	
	M	*SD*	*M*	*SD*
Correct words	13.46	2.10	13.90	1.61
Correct pictures	14.69	1.35	14.28	1.05
Words misattributed to pictures	2.73	2.46	2.10	1.61
Pictures misattributed to words	1.27	1.34	1.90	1.34

Table 5 Imagery results for HPVH and LPVH Groups

	High Group *n* = 26		Low Group *n* = 21	
	M	SD	M	SD
SUIS	41.64	8.72	36.62	8.15
VVIQ	34.35	9.71	39.90	13.23
PIT positive vividness	37.15	4.00	37.04	5.84
PIT negative vividness	34.23	5.73	30.19	7.62
PIT positive likelihood	37.07	4.84	37.48	4.61
PIT Negative likelihood	28.03	6.11	24.52	5.94
PIT positive experiencing	32.38	5.14	31.81	7.87
PIT negative experiencing	27.46	7.56	23.24	6.74

Note. SUIS: spontaneous use of imagery score
VVIQ: vividness of visual imagery questionnaire score, (higher scores indicate lower imagery)
PIT negative: Prospective use of imagery- negative scenario scale, rating scales score vividness, likelihood and experiencing
PIT positive: Prospective use of imagery- negative scenario scale, rating scales score vividness, likelihood and experiencing