Open access • Journal Article • DOI:10.1007/S40732-016-0196-X

A Pilot Study of the Relations Within Which Hearing Voices Participates: Towards A Functional Distinction Between Voice Hearers and Controls — Source link

Ciara McEnteggart, Yvonne Barnes-Holmes, Jos I. M. Egger, Dermot Barnes-Holmes

Institutions: Ghent University, Radboud University Nijmegen

Published on: 17 Aug 2016 - Psychological Record (Springer International Publishing)

A Pilot Study of the Relations within which Hearing Voices Participates:

Towards A Functional Distinction between Voice Hearers and Controls

Ciara McEnteggart[1], Yvonne Barnes-Holmes[1], Jos Egger[2, 3, 4], and Dermot Barnes-Holmes[1]

(1) Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium

(2) Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands

(3) Centre of Excellence for Neuropsychiatry, Vincent van Gogh Institute for Psychiatry, Venray, The Netherlands

(4) Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands

Corresponding author: Ciara McEnteggart

Email: ciara.mcenteggart@ugent.be

Acknowledgements

The data for the current manuscript was collected at the National University of Ireland, Maynooth. It was prepared for publication with the support of an FWO Type I Odysseus Award (2015-2020) at Ghent University, Belgium.

Abstract

The current research used the Implicit Relational Assessment Procedure (IRAP) as a preliminary step toward bringing a broad, functional approach to understanding psychosis, by focusing on the specific phenomenon of auditory hallucinations of voices and sounds (often referred to as hearing voices). On this path, we created a taxonomy of some critical features of voice hearing based on the existing literature (i.e., perceived normality of voices, appraisals of self and other people hearing voices, and fear of voices) as a focus of our experimental manipulations. It was our hope that our findings would add to the broad literature that has used explicit measures to study these phenomena, and that the use of an 'implicit' measure might assist toward a functional-analytic understanding. Three pilot studies were conducted to assess the relations within which hearing voices participates in non-clinical voice hearers (i.e., individuals who hear voices but have no clinical diagnosis or distress) and compared to non-voice hearing control participants. The IRAP effects demonstrated *both* positive and negative relational responses across the three studies, and these effects varied according to explicit levels of delusional ideation. Furthermore, these IRAP effects also predicted explicit aspects of voice hearing and well-being. The current set of pilot studies demonstrate the utility and precision of the IRAP in this domain, and we propose that this type of experimental analysis may hold potential for future bottom-up functional analyses of voice hearing.

Keywords: Implicit Relational Assessment Procedure, psychosis, hearing voices, functional analysis

The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006) is based on Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), a modern functional-analytic approach to human language and cognition. For RFT, verbal histories are established by exemplar training and natural language interactions, through which individuals learn to relate stimuli and events never before related together, and can do so indirectly from specific histories of relating stimuli in similar ways. From a measurement perspective, RFT is interested in targeting these relational responses 'in flight' and exploring the types of verbal histories that give rise to specific verbal repertoires, such as those involved in human psychological suffering (Barnes-Holmes, Barnes-Holmes, Hussey, & Luciano, 2016; Finn, Barnes-Holmes, Hussey, & Graddy, 2016). The IRAP was designed specifically for this purpose, and currently has over 50 published empirical articles supporting its utility (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). Indeed, the IRAP shares many methodological features with implicit measures such as the Implicit Association Test (IAT; e.g., Greenwald, Nosek, & Benaji, 2003). For example, both are automated reaction-time based, group measures in which participants' pair stimuli on a computer screen, and the basic assumption is that participants respond more quickly to stimulus pairings that are consistent (e.g., flowers-pretty and insects-ugly) with their pre-experimental verbal histories than those that are inconsistent (e.g., flowers-ugly and insects-pretty) when they are asked to respond quickly on consistent and inconsistent blocks of trials. On the IRAP, the standardized difference scores between response latencies on consistent and inconsistent blocks of trials generate four D_{IRAP} scores, one for each trial-type (e.g., pleasant-pleasant, pleasant-unpleasant, unpleasant-unpleasant, and unpleasant-pleasant). The IRAP has also demonstrated good reliability and predictive validity (Carpenter, Martinez, Vadhan, Barnes-Holmes, & Nunes, 2013; Fischer, 2013; Vahey, Nicholson, & Barnes-Holmes, 2015).

Using the IRAP to Study Clinical Phenomena

There is an increasing focus on the utility of the IRAP among researchers of clinically-relevant phenomena (see Vahey et al., 2015, for a meta-analysis). One significant advantage of the IRAP over the IAT is its ability to provide greater precision in terms of understanding the observed patterns of relational responding. That is, while the IAT identifies *associations* between pairings, the IRAP also specifies *the nature of* these pairings. Consider the study by Nicholson and Barnes-Holmes (2012a) that presented two IRAPs, both of which assessed disgust toward pleasant pictures (e.g., neatly folded towels) or unpleasant pictures (e.g., a dirty toilet). Specifically, one IRAP assessed disgust propensity (i.e., the tendency to experience disgust), while the other assessed disgust sensitivity (i.e., how negatively a disgust experience is appraised). In simple terms, the disgust propensity IRAP measured *emotional* reactions, while the sensitivity IRAP measured *behavioral* reactions. Participants also undertook a series of behavioral approach tasks (BATs) and explicit measures. The results demonstrated that while responding on both IRAPs predicted obsessive compulsive tendencies on explicit measures of obsessive compulsive disorder (OCD), only the sensitivity IRAP predicted avoidance behaviors on the BAT. That is to say that the behavioral reaction to the event predicted *actual* behavior, while the emotional reaction to it did not. The IRAP's ability to separate these two constructs of the same overarching feature (disgust) highlighted the potential promise of the measure in clinical domains. Furthermore, the IRAP has recently been shown to predict treatment outcomes with cocaine dependence and correctly classify individuals with suicidal ideation from those without (Carpenter et al., 2013; Hussey, Barnes-Holmes, & Booth, 2016).

Hearing Voices Research

Hearing voices is highly prevalent (approx. 70%) in individuals with a diagnosis of schizophrenia (Sartorius et al., 1986). On balance however, it is also prevalent in other psychiatric diagnoses, such as: borderline personality disorder (approx. 32%; e.g., Slotema et

al., 2012); dissociative disorder (approx. 70-90%; e.g., Dorahy et al., 2009); post-traumatic stress disorder (PTSD; approx. 50%); bipolar disorder (approx. 7%; e.g., Blakemore, Smith, Steel, Johnstone, & Frith, 2000) and major depression. Critically, hearing voices is *also* common (approx. 10-15%) in individuals with *no* clinical diagnosis, social and/or occupational dysfunction or psychological distress (Beavan, Read, & Cartwright, 2011; Eaton, Romanoski, Anthony, & Nestadt, 1991; Rössler et al., 2007; Sommer et al., 2010; Tien, 1991).

Despite a substantive body of research using implicit measures in other clinical domains, there appear to be only a handful of published IAT studies and only one IRAP study in the context of psychotic experiences (McEnteggart, Barnes-Holmes, & Adekuoroye, in press). In this study, McEnteggart and colleagues used the IRAP to investigate whether fear of voices would decrease after a hearing voices simulation in a group of non-voice hearers, but interestingly it was found that fear increased. Consistent with the psychosis literature generally, all other studies of voice hearing have relied *largely* on explicit measures (for both clinical and research purposes, see Kim et al., 2010; Ratcliff, Farhall, & Shawyer, 2010). Explicit measures of voice hearing primarily focus on the phenomenological features of voices, appraisals, or reactions to voices. While these studies may seem limited because they relied entirely on explicit measures, it is important to recognize that this type of research has played a key role in understanding and assessing the central features of the voice hearing experience in both clinical and non-clinical populations. Investigations of voice hearing in non-clinical populations are also potentially important. First, it is likely to be the same process through which voices develop in both populations. Second, there may be differences in the ways in which these individuals *respond* to their voices (e.g., appraisals of voices, levels of perceived control, emotional reactions to voices, behavioral reactions to voices, see

Johns et al., 2014 for a review). The latter is important because responding to voices reliably predicts voice-related distress.

An important difference that has emerged in the rapidly-growing research area on appraisals of voices contrasts how benevolence, malevolence and omnipotence differentially influence behavioral responses to voices. In fact, benevolent appraisals have been associated with voice engagement, omnipotent and fearful appraisals with distress, and malevolent appraisals with voice resistance, and each of these are independent of frequency, severity, and intensity of voices (Chadwick & Birchwood, 1994; Gauntlett-Gilbert & Kuipers, 2005; Jackson, Hayward, & Cooke, 2011; Mawson, Berry, Murray, & Hayward, 2011; O'Brien & Johns, 2013; Peters, Williams, Cooke, & Kuipers, 2012). On a broader level, various studies have suggested that mental health labels such as psychosis (as an indication of 'abnormality') can facilitate negative appraisals of voices, especially when the experiences are perceived as rare or unusual (Corrigan, 2004; Mak, Poon, Pun, & Cheung, 2007). Indeed, Corrigan (2004) proposed that self-evaluations must be considered when investigating appraisals of voices among voice hearers, in terms of stigma against the "self" hearing voices *and* against "others" who hear voices.

A Functional Approach

Although almost all schools of thought in psychology have offered comprehensive, eloquent and often overlapping accounts of psychotic experiences, including voice hearing, very little has emerged from the functional-analytic community. For us, this approach would potentially offer an understanding of why and how voice hearing occurs (i.e., what are the key variables and processes involved) and is maintained (i.e., the psychological functions served by these behaviors). In order to make the first step towards a functional-analytic conceptual analysis of voice hearing, we must begin empirically. So, in response to this gap in the literature, the current set of pilot studies sought to determine the utility of the IRAP, as

a complement to traditional explicit measures. We opted to focus specifically on the IRAP because of the level of concept precision it had demonstrated previously in the context of clinical domains. Although the IRAP is a group measure, which is not typically employed in traditional single subject functional analysis, it appears that although individual patterns of behavior are unique, they do not vary widely from each other at the group level.

Numerous studies have examined the predictive utility of the IRAP (Carpenter et al., 2013) and various other features (e.g., test-retest reliability, see Fischer, 2013; Vahey et al., 2015). The findings from these and the 50+ published IRAP studies lead us to conclude that the IRAP is a sound and precise measure of relational responding – that is what the measure does. Specifically, it presents relations on-screen (usually coordination versus distinction) and assesses the speed, and ways in which, participants engage in the target relational responses. While numerous studies have increasingly used the IRAP to explore clinical phenomena, it remains the case that even here, the measure simply assesses the accuracy and speed at which participants derive the relations presented on-screen. Of course, those studies become more domain-relevant because they specifically seek to determine whether these relational responses predict scores on standardized explicit measures pertinent to the target domain. However, such studies are often preceded by preliminary experimental work to identify which relational responses seem most pertinent in a given domain. The IRAP work on OCD is a prime example. The current work is of the preliminary variety in this regard. We examined the literature carefully and tried to determine which relational responses appear to underpin the types of phenomena referred to in the literature. We then targeted those relations in various IRAPs and presented them to samples who might differ in this regard. Ultimately, our aim is to build a research program that will highlight the potentially different functions of voice hearing relations among these groups.

The current set of studies did not assess responding to voices directly, as traditionally defined. From an RFT perspective, the studies assessed the types of relations within which voices participated. In Study 1, for instance, the IRAP presented participants with opportunities to coordinate or distinguish 'hearing voices' with normality and abnormality (e.g., Hearing voices is-Normal-True indicates a coordination relation between voice hearing and normality). In Study 2, the IRAP presented opportunities to coordinate or distinguish 'hearing voices' with fear and acceptability (e.g., Hearing voices is-Scary-False indicates a distinction relation between hearing voices and fear). And finally, the IRAP in Study 3 presented opportunities to coordinate or distinguish 'my hearing voices' and 'others hearing voices' with fear and acceptability (e.g., If I heard voices-It would be scary-False indicates a distinction relation between my hearing voices and fear).

We investigated these relational responses to voices with non-clinical voice hearers and non-voice hearing controls. While the existing literature shows topographical similarities between hearing voices in clinical and non-clinical populations, there have been little or no studies comparing non-clinical voice hearers with samples who do not appear to hear voices. In examining this comparison, we wanted to get a sense of the relations within which voice hearing participated for non-clinical voice hearers, and how this might compare with controls. It is important to note that we do not assume generalization of our predictions or findings to clinical voice hearers. In fact, if voice hearing participates in different relations for these two groups, one would predict the transformation of different functions (McEnteggart et al., under review). This type of further clinical research clearly required baseline observations of the relevant relational responses of non-voice hearers. This was the aim of the current pilot work. Ultimately, any differences we might observe in the relations within which voice hearing participates for the two groups might inform future comparisons between clinical and non-clinical voice hearers and might shed light on potentially different transformations of

functions. This line of work, therefore, may ultimately show differential functions of these experiences for different groups and its potential influence on clinical distress.

Study 1: Assessing the Normality of Hearing Voices

Method

Setting

All participation was on an individual basis. Experimental sessions lasted between 30 and 60 minutes, and all participation was completed in one session. The experimenter interacted with participants only during instructional phases of the IRAP and remained seated behind participants at all other times.

Participants

The current study involved two groups of participants recruited from a general pool of undergraduate students at the National University of Ireland Maynooth. One group was categorized as non-clinical voice hearers and the other group comprised a non-voice hearing control group. Seven non-clinical voice hearers were identified as such using Item No. 33 of the Community Assessment of Psychic Experiences (CAPE). Thirty-six non-voice hearing individuals were identified. In total, the study involved 43 participants, with an age range of 18 to 38 years and a mean of 22.16 years.

Materials

Explicit measures. Two broad categories of explicit measures were administered. The first series of measures assessed voice hearing (the Auditory Hallucinations Rating Scales, the Beliefs about Voices Questionnaire-Revised, the Voices Acceptance and Action Questionnaire) and delusional ideation (the CAPE). These measures are widely used in the assessment of voice hearing (see Ratcliff, Farhall, & Shawyer, 2010), and the aim of their inclusion was to examine the predictive validity of the IRAP data in this domain, and not in

the diagnosis of a psychotic disorder. In particular, the CAPE is used extensively to assess the presence of hearing voices and it was used for this purpose here.

Community Assessment of Psychic Experience (CAPE). The CAPE is a 42-item measure of delusional ideation (or psychotic experiences) in the general population (Stefanis et al., 2002). The measure assesses three dimensions of psychotic symptoms: positive (e.g., hallucinations or delusions), negative (e.g., social withdrawal), and depressive. All items are rated in terms of frequency on a 4-point scale from 0 (never) to 1 (nearly always) and similarly rated in terms of level of distress from 0 (not distressed) to 3 (very distressed). These two sets of ratings on each dimension yield six independent weighted scores (i.e., two scores for each dimension) that indicate high or low frequency or distress on each dimension (with a maximum score of 6.0 on each). This scale has demonstrated adequate reliability with an alpha coefficient of 0.63 for the positive dimension, 0.64 for the negative dimension and 0.62 for the depressive dimension (Konings, Bak, Hanssen, Van Os, & Krabbendam, 2006).

Auditory Hallucinations Ratings Scale (AHRS). The AHRS is a subscale of the Psychotic Symptom Rating Scales (PSYRATS) and excludes an additional subscale that measures delusions (Haddock, McCarron, Tarrier, & Farragher, 1999). The AHRS is an 11-item scale that assesses the severity of 11 target dimensions of voice hearing (e.g., degree of negative content - minority of voice content is unpleasant or negative). All items are rated on a 5-point scale from 0 (e.g., voices not present) to 4 (e.g., voices present most of the time). The AHRS yields an overall score with a maximum of 44 indicating high degrees of voice hearing and a minimum of 0 indicating low degrees of voice hearing. This scale has demonstrated excellent inter-rater reliability (ICC = 0.99 - 1.00) and test-retest reliability (ICC = 0.70; Drake, Haddock, Tarrier, Bentall, & Lewis, 2007; Haddock et al., 1999).

Beliefs About Voices Questionnaire–Revised (BAVQ-R). The BAVQ-R is a 35-item scale that targets beliefs, feelings and behaviors about voice hearing (Chadwick, Lees, &

Birchwood, 2000). The measure comprises seven subscales: malevolence; benevolence; omnipotence; emotional resistance; behavioral resistance; emotional engagement; and behavioral engagement. All items are rated on a 4-point scale from 0 (disagree) to 3 (strongly agree). Subscales are scored independently and indicate high or low levels of each dimension. The BAVQ-R yields a minimum score of 0 for all subscales, and a maximum score of: 18 for malevolence, benevolence and omnipotence; 12 for emotional resistance, emotional engagement and behavioral engagement; and 15 for behavioral resistance. The BAVQ-R subscales have demonstrated adequate internal consistency with an alpha coefficient of = 0.74 to 0.88 (Chadwick et al.).

Voices Acceptance and Action Questionnaire (VAAS). The VAAS is a 31-item scale that measures acceptance of voices (Shawyer et al., 2007). The scale comprises two broad sections that measure emotional acceptance and behavioral acceptance. All items are rated on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree). The VAAS yields an overall score with a maximum of 155 indicating high voice acceptance and a minimum of 0 indicating low acceptance. This scale has demonstrated excellent internal consistency with an alpha coefficient of 0.90 (Shawyer et al.).

The Acceptance and Action Questionnaire II (AAQ-II) more broadly assessed general psychological well-being and was included to control for subclinical levels of distress in the control participants and to ensure that the sample of voice hearers did not contain a mix of those who were or were not clinically distressed.

Acceptance and Action Questionnaire II (AAQ-II). The AAQ-II is a 10-item measure of psychological inflexibility around negative private events (Bond et al., 2011). All items are rated on a 7-point scale from 1 (never true) to 7 (always true). The AAQ yields an overall score with a maximum of 70 indicating *high* psychological inflexibility and a minimum of 10 indicating *low* psychological inflexibility. This scale has demonstrated

adequate internal consistency with an alpha coefficient of 0.78 to 0.88 across several samples (Bond et al.).

The IRAP. The IRAP was administered on a standard desktop computer. This was used to present the instructions and stimuli, and to record responses. The current study involved one IRAP that assessed relational responses to hearing voices and normality/abnormality, hereafter referred to as the Normality IRAP. The IRAP contrasted hearing voices with seeing things, using the labels HEARING VOICES IS and SEEING THINGS IS. Each trial-type presented one of these two types of category labels. These were accompanied by one of three positive (e.g., NORMAL) or three negative target stimuli (e.g., ABNORMAL). The screen also presented two response options, TRUE and FALSE. Before each block of trials, the screen presented one of two rules for responding (i.e., PLEASE ANSWER AS IF HEARING VOICES IS ABNORMAL AND SEEING THINGS IS NORMAL or PLEASE ANSWER AS IF HEARING VOICES IS NORMAL AND SEEING THINGS IS ABNORMAL). A full list of label stimuli, target stimuli, and response options for the IRAP is provided in Table 1.

INSERT TABLE 1 HERE

Procedure

The current study comprised two stages, one involving the explicit measures and the second involving the IRAP. All participants were instructed that experience of hearing voices was the focus of the study. However, in order to ensure that the study was accurately measuring appraisals to voice hearing as "auditory verbal hallucinations" and no other phenomena, all participants were provided with a written explanation of voice hearing, and instructed that this was the focus of the study.

Explicit Measures. Participants were then identified as either non-clinical voice hearers or non-voice hearing controls using the current screening measures. Specifically, if

participants indicated on the CAPE that they did not hear voices, they were allocated to the non-voice hearing control group. These individuals were *not* thereafter presented with the other two measures of voice hearing because these are constructed in such a way that they assume respondents hear voices and make little sense to individuals who do not. In contrast, if participants indicated on the CAPE that they *did* hear voices, they were allocated to the voice hearing group and then presented with the other three measures of voice hearing.

All participants were presented with the AAQ-II. We then used scores on the AAQ-II to ensure that those who had identified themselves on the CAPE as voice hearers did not indicate clinical distress on the AAQ-II. Based on the AAQ scores, we then categorized all voice hearers as non-clinical and aimed to exclude voice hearers who showed clinical distress on the AAQ from further participation.

IRAP. The verbal and automated instructions provided to participants for completing each IRAP were consistent with those in the most recently published IRAP research (e.g., Nicholson & Barnes-Holmes, 2012b). In short, there were three key features of the task: the criterion for high levels of accurate responding (i.e., 80%), the criterion for responding very quickly (i.e., <2,000 ms.), and the fact that the 'correct' and 'incorrect' patterns of responding (depending upon the rule provided) alternated across blocks of trials. Similarly, the presentation features of the IRAP were identical to most recently published work in terms of: a maximum of four pairs of practice blocks (depending upon performance); three pairs of test blocks; 24 trials in every block; four trial-types; and two specified rules for responding. For all participants, blocks alternated between two patterns of responding according to the specified rules (e.g., responding as if voices are normal vs. responding as if voices are abnormal, see Figure 1). Blocks were counterbalanced across participants in terms of which rule was presented first (e.g., Rule A in the first block, Rule B in the second block, Rule A in

the third block and so on). For illustrative purposes, see Figure 1 for a schematic representation of the screen presentation of the Normality IRAP.

INSERT FIGURE 1 HERE

Fig 1. Schematic representation of the four trial-types presented in the Normality IRAP. The arrows and text boxes did not appear on the participant's screen, they indicate the correct responses for Rule A and Rule B blocks of trials. The labels used for the four trial-types are as follows: Voices-Abnormal (top-left), Voices-Normal (top-right), Visions-Abnormal (bottom-left) and Visions-Normal (bottom-right).

Once participants finished the IRAP, they completed the five explicit measures in a pre-determined sequence (AHRS, BAVQ-R, VAAS, CAPE, and AAQ).

Results

Explicit Measures Data

The means and standard deviations were calculated from each participant's responses on each of the explicit measures and data are summarized in Table 2.

INSERT TABLE 2 HERE

IRAP Data

Scoring of the IRAP was always conducted using the standardized algorithm for transforming the difference in latencies between consistent and inconsistent blocks of trials into D_{IRAP} scores (see Nicholson et al., 2013). All data from any participant that fell below 80% accuracy and above 2000ms latency on any of the six test blocks were omitted from the dataset (N=7). The final dataset comprised 36 participants: 29 non-voice hearers (15=male, 12=female); and seven non-clinical voice hearers (4=male, 3=female).

Analyses included between group analyses, delusional ideation analyses, and correlational analyses. Delusional ideation analyses categorized the non-voice hearing participants according to their positive psychotic symptom scores on the CAPE, and aimed to

investigate whether higher levels of delusional ideation on this subscale may be associated with IRAP effects that are comparable to those of the voice hearers.

Between groups analyses (non-clinical voice hearers and non-voice hearers). The mean D_{IRAP} scores for the IRAP are presented in Figure 1. On the Voices-Normal trial-type, both groups showed voices-normal effects (i.e., participants responded faster on Voices-Normal-True). A similar pattern emerged on the Voices-Abnormal trial-type. However, on both trial-types, non-clinical voice hearers had greater effects. In order to investigate the effects of group on trial-type, analyses of variance found a main effect for trial-type (df=1, F=8.935, p<0.01, μ^2=0.842). Post-hoc analyses revealed no significance between the two groups on each trial-type (*all p's*>0.05). Furthermore, one-sample t-tests investigated whether the D_{IRAP} effects differed significantly from zero and found that for both groups, Voices-Normal was significant (voice hearers: df=6, t=3.639, p<0.01; non-voice hearers: df=28, t=3.811, p<0.001).

INSERT FIGURE 2 HERE

Fig 2. Mean D_{IRAP} scores for the two groups on the Normality IRAP. Positive D_{IRAP} scores indicate voices-normal effects and negative D_{IRAP} scores indicate voices-abnormal effects. Significant D_{IRAP} effects are denoted by *.

Delusional ideation analyses. For these analyses, data from the non-voice hearers was split into two groups using a median split on the positive dimension subscale, thus creating high versus low non-voice hearing comparison groups. These were then also compared with the voice hearers, all of whom scored high on the CAPE (M=2.45, SD=0.48). These were then also compared with the voice hearers, all of whom scored high on the CAPE. The mean D_{IRAP} scores for voice hearers (N=7) and high (N=16, M=2.391, SD=0.53) and low (N=13, M=1.52, SD=0.21) positive dimension scores (non-voice hearers) on the IRAP are presented in Figure 3.

For Voices-Normal, the voice hearers showed the greatest voices-normal effects, followed by the high CAPE group and then the low CAPE group. For Voices-Abnormal, the voice hearers again showed voices-normal effects (i.e., participants responded faster to Voices-Abnormal-False), followed by the high CAPE group. The low CAPE group showed marginal voices-abnormal effects (i.e., participants responded faster to Voices-Abnormal-True). Analyses of variance showed a significant main effect for trial-type (F=10.774, p<0.01, μ^2=0.907). Again, one-sample t-tests indicated that for all three groups, significant D_{IRAP} effects were found for Voices-Normal (low: df=12, t=2.829, p<0.01; high: df=15, t=2.688, p<0.01; voice hearers: df=6, t=3.639, p<0.01).

INSERT FIGURE 3 HERE

Fig 3. Mean D_{IRAP} scores on the Normality IRAP for voice hearers and high and low CAPE groups scores. Positive D_{IRAP} scores indicate voices-normal effects and negative D_{IRAP} scores indicate voices-abnormal effects. Significant D_{IRAP} effects are denoted by *.

Correlations

A correlation matrix was conducted between the IRAP and the explicit measures using the voice hearers' data only, because the primary aim was to investigate if IRAP effects correlated with explicit measures of voice hearing. For Voices-Normal, a significant positive correlation was found with behavioral engagement on the BAVQ-R (r=0.336, p<0.05) and voice acceptance (r=0.319, p<0.05). That is, the greater the behavioral engagement with voices and acceptance of voices, the greater the voices-normal effect.

For Voices-Abnormal, a significant positive correlation was found with BAVQ-R benevolence appraisals (r=0.331, p<0.05) and emotional engagement with voices (r=0.356, p<0.05), with CAPE positive distress (r=0.332, p<0.05), the positive dimension (r=0.333, p<0.05), the depressive frequency (r=0.331 p<0.05), negative distress (r=0.352, p<0.05), negative frequency (r=0.452, p<0.01), and CAPE negative dimension (r=0.408, p<0.01). That is, the greater voices-abnormal effects on Voices-Abnormal, the greater: benevolence

appraisals, emotional engagement with voices, positive, negative and depressive delusional experiences.

Study 2: Assessing the Fear of Hearing Voices

Method

Setting

All aspects of the experimental setting were identical to Study 1.

Participants

The current study involved two groups of participants from a general sample of undergraduate students. One group was categorized as non-clinical voice hearers and the other group comprised a non-voice hearing control group. Seventeen non-clinical voice hearers and 25 non-voice hearing individuals were identified as such using current screening methods from a general sample of undergraduate students. The age range of participants was 18 to 37 years, with a mean age of 22.18 years.

Materials

Explicit measures. All explicit measures were identical to Study 1.

The IRAP. The current study involved one IRAP that assessed relational responses to hearing voices and fear/acceptability (referred to as the Fear IRAP).

The IRAP contrasted hearing voices with seeing things, using the labels HEARING VOICES IS and SEEING THINGS IS. Each trial-type presented one of these two types of category labels. These were accompanied by one of three positive (e.g., FINE) or three negative target stimuli (e.g., SCARY). The screen also presented two response options, TRUE and FALSE. Before each block of trials, the screen presented one of two rules for responding (i.e., PLEASE ANSWER AS IF HEARING VOICES IS SCARY AND SEEING THINGS IS OKAY or PLEASE ANSWER AS IF HEARING VOICES IS OKAY AND

SEEING THINGS IS SCARY). A full list of label stimuli, target stimuli, and response options for the IRAP is provided in Table 3.

INSERT TABLE 3 HERE

Procedure

All aspects of the experimental procedure were identical to Study 1. For illustrative purposes, see Figure 4 for a schematic representation of the screen presentation of the IRAP.

INSERT FIGURE 4 HERE

Fig 4. Schematic representation of the four trial-types presented in the Fear IRAP. The arrows and text boxes did not appear on the participant's screen, they indicate the correct responses for Rule A and Rule B blocks of trials. The labels used for the four trial-types are as follows: Voices-Scary (top-left), Voices-Okay (top-right), Visions-Scary (bottom-left) and Visions-Okay (bottom-right).

Results

Explicit Measures Data

The means and standard deviations were calculated from each participant's responses on each of the explicit measures and data are summarized in Table 4.

INSERT TABLE 4 HERE

IRAP Data

Scoring of the IRAP was identical to Study 1. Data from any participant that fell below 80% accuracy and above 2000ms latency on any of the six test blocks was omitted from the dataset (N=15). The final dataset comprised 34 participants: 17 non-voice hearers (10 male and 7 female) and 17 non-clinical voice hearers (9 male and 8 female).

Between groups analyses. The mean D_{IRAP} scores for the three groups on the IRAP are presented in Figure 5 (visions trial-types are again excluded). On Voices-Okay, the controls and non-clinical voice hearers showed pro-voices effects. On the Voices-Scary trial-type, both groups showed anti-voices effects. Analyses of variance revealed no significant

main effects (all *p's*>0.05), however there was an interaction effect between group and trial-type (F=3.660, p<0.05, $\mu 2$=0.631). Post-hoc analyses revealed no differences between the groups on each trial-type (all *p's*>0.05). Indeed, the D_{IRAP} trial-type effect was significant for the non-voice hearers on Voices-Okay (df=16, t=-3.010, p<0.01, all other *p's*>0.05).

INSERT FIGURE 5 HERE

Fig 5. Mean D_{IRAP} scores for the two groups on the Fear IRAP. Positive D_{IRAP} scores indicate pro-voices effects and negative D_{IRAP} scores indicate anti-voices effects. Significant D_{IRAP} effects are denoted by *.

Delusional ideation analyses. Data from the non-voice hearers was again split into two groups using a median split on the positive dimension subscale, thus creating high vs. low CAPE non-voice hearing comparison groups. These were then also compared with the voice hearers, all of whom scored high on the CAPE. The mean D_{IRAP} scores for voice hearers (N=17, M=2.23, SD=0.84) and high (N=8, M=2.52, SD=0.37) and low (N=9, M=1.64, SD=0.37) positive dimension CAPE scores (non-voice hearers) on the IRAP are presented in Figure 6. For Voices-Okay, all three groups showed pro-voices effects, with the high CAPE group showing the largest. For Voices-Scary, all three groups showed comparable anti-voices effects. Analyses of variance produced a main effect for trial-type (F=3.317, p<0.0001, μ^2=0.998). Exploratory trial-type analyses found no significant differences between the groups (all *p's*>0.05). A significant D_{IRAP} effect was only found for Voices-Okay for the high CAPE group (df=7, t=-2.403, p<0.05).

INSERT FIGURE 6 HERE

Fig 6. Mean D_{IRAP} scores on the Fear IRAP for voice hearers and high and low CAPE groups on the Fear IRAP. Again, positive D_{IRAP} scores indicate pro-voices effects and negative D_{IRAP} scores indicate anti-voices effects. Significant D_{IRAP} effects are denoted by *.

Correlations

A correlation matrix also investigated potential relationships between the IRAP effects and the explicit measures among the voice hearers, but no significant correlations were found (all *p's*>0.05).

Study 3: Appraisals of Self and Others Hearing Voices

Method

Setting

All aspects of the setting in Study 3 were identical to the two previous studies.

Participants

From a general sample of undergraduate students, the current study involved two groups of participants. One group was categorized as non-clinical voice hearers and the other group comprised a non-voice hearing control group. Twenty-four non-clinical voice hearers and 43 non-voice hearing individuals were identified using current screening methods. In total, the study involved 67 participants, with an age range of 19 to 38 years and a mean of 23.8 years.

Materials

Explicit measures. All explicit measures were identical to the previous two studies.

The IRAP. The current study involved two IRAPs, one that assessed relational responses regarding the *self* hearing voices (referred to as the Self IRAP) and *other people* hearing voices (referred to as the Others IRAP) in the context of fear/acceptability.

The Self IRAP contrasted hearing voices with seeing things, using the labels IF I HEARD VOICES and IF I SAW THINGS. Each trial-type presented one of these two category labels, accompanied by one of three positive (e.g., IT WOULD BE FINE) or three negative target stimuli (e.g., IT WOULD BE SCARY). The screen also presented the response options TRUE and FALSE. Before each block of trials, the screen presented one of

two rules for responding (i.e., PLEASE ANSWER AS IF HEARING VOICES IS OKAY AND SEEING THINGS IS SCARY or PLEASE ANSWER AS IF HEARING VOICES IS SCARY AND SEEING THINGS IS OKAY).

The Others IRAP contrasted hearing voices with seeing things, using the labels IF OTHER PEOPLE HEARD VOICES and IF OTHER PEOPLE SAW THINGS. Each trial-type presented one of these two category labels, accompanied by one of three positive (e.g., IT WOULD BE FINE) or three negative target stimuli (e.g., IT WOULD BE SCARY). The screen also presented the response options TRUE and FALSE. Before each block of trials, the screen presented one of two rules for responding (i.e., PLEASE ANSWER AS IF HEARING VOICES IS OKAY AND SEEING THINGS IS SCARY or PLEASE ANSWER AS IF HEARING VOICES IS SCARY AND SEEING THINGS IS OKAY). A full list of label stimuli, target stimuli, and response options for the IRAP is provided in Table 5.

INSERT TABLE 5 HERE

Procedure

The current study comprised two IRAPs: The Self IRAP; and The Others IRAP. For all participants, there were two stages, one involving the IRAP and the second involving the explicit measures. It is important to emphasize, therefore, that each participant only completed *one IRAP*, the Self IRAP *or* the Others IRAP, hence approximately one half of each group of participants completed each IRAP (i.e., half of the non-voice hearing control group completed the Self IRAP, while the other half completed the Others IRAP, and similarly half of the non-clinical voice hearing group completed the Self IRAP, while the other half completed the Others IRAP). Participants were randomly assigned to either IRAP. For illustrative purposes, see Figure 7 for a schematic representation of the screen presentation of the IRAPs.

INSERT FIGURE 7 HERE

Fig 7. Schematic representation of the four trial-types presented in the Self (left) and Others (right) IRAPs. The arrows and text boxes did not appear on the participant's screen, they indicate the correct responses for Rule A and Rule B blocks of trials. The labels used for the four trial-types in each IRAP are as follows: Voices-Scary (top-left), Voices-Okay (top-right), Visions-Scary (bottom-left) and Visions-Okay (bottom-right).

All other aspects of the experimental procedure were identical to the previous two studies.

Results

Explicit Measures Data

The means and standard deviations were calculated from each participant's responses on each of the explicit measures and data are summarized in Table 6.

INSERT TABLE 6 HERE

IRAP Data

Scoring of the IRAP was identical to the previous two studies. All data from any participant that fell below 80% accuracy and above 2000ms latency on any of the six test blocks were omitted from the dataset (N=19). The final dataset comprised 48 participants (22=male, 26=female): 23 non-voice hearers (13 in Self IRAP and 10 in Others IRAP); and 25 non-clinical voice hearers (14 in the Self IRAP and 11 in the Others IRAP).

Between groups analyses. The mean D_{IRAP} scores for the IRAP are presented in Figure 8 (visions trial-types are again excluded). On the Self IRAP, on Voices-Okay, both groups showed a similar pro-voices effect. On Voices-Scary, both groups showed anti-voices effects, although the voice hearers' effect was negligible. On the Others IRAP, on Voices-Okay, both groups showed pro-voices effects, whereas on Voices-Scary, both groups showed anti-voices effects. Analyses of variance found a main effect for trial-type (df=1, F=25.884, p<0.0001, μ^2=1.0), and an interaction effect between trial-type, group and IRAP (df=1, F=4.361, p<0.05, μ^2=0.522). Post-hoc analyses found differences between the groups on the Self IRAP for Voices-Scary (df=25, t=-2.107, p<0.05). Further analyses investigated

potential differences between the IRAPs for each group, but found none (*all p's*>0.05). For the controls, both D_{IRAP} effects on the Self IRAP were significant (Voices-Scary: *df*=12, *t*=3.277, *p*<0.01; Voices-Positive: *df*=12, *t*=-2.442, *p*=0.05).

INSERT FIGURE 8 HERE

Fig 8. Mean D_{IRAP} scores for the two groups on the Self and Others IRAPs. Positive D_{IRAP} scores indicate pro-voices effects and negative D_{IRAP} scores indicate anti-voices effects. Significant D_{IRAP} effects are denoted by *.

Delusional ideation analyses. Data from the non-voice hearers was again split into two groups using a median split on the positive dimension subscale, thus creating high vs. low CAPE non-voice hearing comparison groups. Once again, these were then compared with the voice hearers, all of whom scored high on the CAPE. The mean D_{IRAP} scores for voice hearers (self: *N*=14, *M*=3.04, *SD*=1.60, others: *N*=11, *M*=2.47, *SD*=0.72) and high (self: *N*=4, *M*=2.48, *SD*=0.62, others: *N*=9, *M*=2.30, *SD*=0.51) and low (self: *N*=4, *M*=1.52, *SD*=0.16, others: *N*=6, *M*=1.64, *SD*=0.20) positive dimension CAPE scores (non-voice hearers) on the IRAP are presented in Figure 9. On the Self IRAP, for Voices-Okay, all three groups showed pro-voices effects, with the high CAPE group showing the largest. For Voices-Scary, the high and low CAPE groups showed anti-voices effects, whereas the voice hearers showed null effects. On the Others IRAP, for Voices-Okay, all three groups showed comparable pro-voices effects. For Voices-Scary, the high CAPE group and the voice hearers showed anti-voices effects, whereas the low CAPE group showed null effects. Analyses of variance produced a main effect for trial-type (*F*=35.215, *p*<0.0001, μ^2=1.0) and an interaction effect for trial-type and CAPE group (*F*=3.891, *p*<0.05, μ^2=0.669). Trial-type analyses only found significant differences between the voice hearers and the high CAPE group on the Self IRAP (*df*=16, *t*=2.241, *p*<0.05). Significant D_{IRAP} effects were only found

on the Self IRAP for the high CAPE group on Voices-Okay (df=3, t=-3.770, p<0.05) and Voices-Scary (df=3, t=-3.770, p<0.05).

INSERT FIGURE 9 HERE

Fig 9. Mean D_{IRAP} scores on the Self and Others IRAPs for voice hearers and high and low CAPE groups on both IRAPs. Again, positive D_{IRAP} scores indicate pro-voices effects and negative D_{IRAP} scores indicate anti-voices effects. Significant D_{IRAP} effects are denoted by *.

Correlations

A correlation matrix also investigated potential relationships between the IRAP effects and the explicit measures among the voice hearers in each IRAP. On the Self IRAP, a significant positive correlation was found between Voices-Okay and Depressive Frequency (CAPE: r=0.622, p<0.05). That is, the greater the depressive frequency the greater acceptability toward voices. In the Others IRAP, a significant negative correlation was found between Voices-Scary and VAAS (r=-0.627, p<0.05), that is, the more the anti-voices effect, the more acceptance of voices.

Discussion

The Current Findings

The current set of preliminary studies were designed to take the first step towards a functional investigation of voice hearing in non-clinical populations. Study 1 used the Normality IRAP to assess relational responses to hearing voices and normality/abnormality. Both non-clinical voice hearers and controls coordinated voice hearing with normality, and this effect was stronger for the voice hearers. This finding appears to contradict existing evidence that some voice hearers categorize voice hearing as *abnormal.* However, it is important to emphasize that the current sample contained only non-voice hearers and non-clinical voice hearers and no clinical voice hearers (Corrigan, 2004; Mak et al., 2007). Interestingly, participants who were high on delusional ideation had the strongest effects for normality, while those who were low were the weakest, and even showed marginal anti-

normality. Indeed, this normality effect predicted behavioral engagement with voices and voice acceptance, while the abnormality effect predicted, benevolent voice appraisals, emotional engagement with voices, and other delusional experiences. These findings are largely consistent with what the literature has recorded with clinical voice hearers (e.g., Chadwick & Birchwood, 1994). It may seem counterintuitive that coordinating voices with abnormality predicted voice benevolence, but even voice hearers who gain guidance and support from voices recognize that these are unusual experiences. Indeed, the aim of the IRAP is to highlight how various relations containing voice hearing can coexist.

Study 2 used the Fear IRAP to assess relational responses to hearing voices and fear/acceptability. Both non-clinical voice hearers and controls coordinated hearing voices with *both* acceptability and fear. Similar to Study 1, participants who were high on delusional ideation had the greatest acceptability effects. Study 3 used the Self and Others IRAPs to assess relational responses to the self and others hearing voices and fear/acceptability. Overall, hearing voices was coordinated with acceptability by both groups on both IRAPs, although control participants were more acceptable on the Self IRAP. Interestingly, control participants were also more fearful on the Self IRAP, while the voice hearers were more fearful on the Others IRAP. In simple terms, voices are more frightening and harder to accept when I have no experience of hearing voices, however, they are more frightening and harder to accept in other people when I have experience of hearing voices. This latter may highlight how voice hearers' stigmatize other voice hearers, despite accepting their own experiences. This finding again shows the independence of these relations as they pertain to the self and others.

Despite the consistency of the IRAP effects across the three studies (i.e., both positive and negative effects were observed), the delusional ideation analyses in Study 3 generated divergent effects. Voice hearers who reported themselves as high on delusional ideation were

most accepting of voices overall and most fearful in the context of self, but voice hearers were the most fearful in the context of others. Correlations revealed that, for the voice hearers, acceptability of voices on the Self IRAP predicted high depressive CAPE symptoms. This may appear contrary to previous evidence that there is an inverse relationship between benevolent appraisals and depression (van der Gaag, Hageman, & Birchwood, 2003). However, acceptability is not directly comparable to benevolent appraisals. Furthermore, fear on the Others IRAP predicted higher voice acceptance, thus showing a clear difference between emotional and behavioral responses to voices. Put simply, you can respond to a voice negatively *while still* showing acceptance towards it.

Overall, some of the findings were consistent with the literature on psychosis (Chadwick & Birchwood, 1994; Corrigan, 2004; Mak et al., 2007; van der Gaag et al., 2003). Yet, these findings also make a noteworthy contribution to the field. For example, we found that non-clinical voice hearers were more positive about voices than one would typically expect, and these mixed beliefs can interact with level of distress (see also Miller, O'Connor, & DiPasquale, 1993; Morrison, Nothard, Bowe, & Wells, 2004).

IRAP Effects

At this stage, we feel it is important to highlight some statistical and interpretive issues that surround IRAP effects, which at a glance may appear to limit our findings and especially our interpretations. As functional-analytic psychologists interested in the key psychological processes in specific patterns of verbal behavior, especially those that contribute to human suffering, it is important to have a reliable means of measuring these processes. Indeed, IRAP research pivots around IRAP effects, although the precise nature and size of these vary considerably across studies (see Hussey, Thompson, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2015). In short, IRAP effects are recorded as any effect that differs from a zero D_{IRAP} score. While at a more stringent technical level, one could argue that

a D_{IRAP} score should differ *significantly* from zero, many IRAP studies (especially clinical ones) cannot hold strictly to this criterion, primarily because of relatively small *N*s. As a result, IRAP researchers often reflect upon the meaning of non-significant IRAP effects, and this is the position in which we found ourselves in the current explorations.

Although some of the effects outlined here were not *statistically* significant from zero, they nonetheless can be *functionally* meaningful. Consider, for example, effects recorded on a Voices-Abnormal trial-type. A D_{IRAP} score of zero would indicate that participants responded to Voices-Abnormal-True and Voices-Abnormal-False at equal speed. For RFT, this suggests that according to an individual's verbal history, there is the same likelihood of deriving one relation over the other, demonstrating relational flexibility for either relation. On the other hand, a D_{IRAP} score which is statistically different from zero suggests a history of relating those stimuli more rigidly, relatively speaking. For example, on Voices-Abnormal, if there was a D_{IRAP} score of -0.5, participants had responded to Voices-Abnormal-True more quickly than Voices-Abnormal-False, because of a history in which there were more opportunities to derive the former. The IRAP thus points toward the key functional processes behind patterns of relational responding and may help when categorizing behavior as, for example, flexible or rigid. Furthermore, it is worth noting that if there was, for example, a voices-abnormal effect on the Voices-Normal trial-type this might suggest, functionally speaking, that participants *rejected normality* toward voices (i.e., participants responded more quickly on Voices-Positive-False), which is different to saying there was an anti-voices effect on this trial-type. We would argue that even those small distinctions may specify different functional patterns of responding.

Implications for the Literature

The IRAP. An aim of the current pilot research was to investigate the potential utility of the IRAP in studying features of psychotic experiences, specifically hearing voices. We

wanted to take the first steps to understanding this phenomenon by taking a very broad approach, so that we can start to look at more complex aspects of the voice hearing experience. The results have demonstrated that the IRAP can potentially deliver a high level of precision that can add to data obtained from traditional explicit measures. For example, IRAP effects predicted several aspects of self-reported voice-related behavior. It is through this precision that we can begin to look at the functional processes (i.e., the key relations) at play in the voice hearing experience.

Voice hearing. The current research also sought to investigate responding to voices as a complement to explicit measures, with a view to obtaining a broader understanding of the onset, maintenance, and the experiential nature of hearing voices in clinical and distressed populations. Specifically, this has begun to help us to investigate very particular features of voice hearing, such as the perceived normality/abnormality, fear/positivity, and acceptability of self and others hearing voices, and their relationship with distress. This shows promise for the level of psychological precision needed to ask complex questions about these experiences, and the IRAP has allowed us to take the first steps towards a more functional understanding of the phenomenon of hearing voices. Moreover, based on these preliminary findings in Study 3 that the non-clinical voice hearers were positive and accepting of their voices and low in clinical distress, the data also speak directly about the types of support or interventions that promote acceptance to reduce distress, and which would be of most benefit to clinical voice hearers.

Functional analysis. As functional contextualists, we naturally began this research with a functional-analytic aim, which not only informs the basic science, but also clinical applications. So, beginning to try to answer functional-analytic questions about voice hearing, as outlined in this pilot set of studies, will hopefully allow us to identify the key processes involved in this behavior, as our overarching aim. Thus, through this research, we have begun

to move towards a more functional-analytic understanding of voice hearing, albeit in a very small way, but crucially in a way that will help guide future studies in this domain. This is done with the hope that these will allow us to better understand these experiences, and perhaps ultimately change them, in the service of the individual.

Limitations

There are a number of limitations of these studies which should be reflected upon when planning future research. First, the time-point at which participants were hearing voices (i.e., past/present) was not controlled and may have influenced the analyses. Second, the current sample comprised a high proportion of psychology undergraduates exposed to some level of psychological training, which may account for some of the normality effects. It would be interesting, therefore, to replicate this study in a sample with no training in psychology. Third, the use of negatively worded target stimuli (i.e., could *not* cope, could *not* accept it) may be difficult for participants to interpret when undertaking the IRAP. Future research should try to circumvent this issue by avoiding the use of 'not' in target phrases. Fourth, the use of inferential statistics and correlational analyses in low *N* analyses does not allow researchers to observe the differences or effects that may exist. Future research should try to include larger *N's* for the analyses in order to better examine the likely nuanced relationships among level of delusional ideation, distress, and IRAP effects. Fifth, the current study was carried out on a non-clinical (i.e., non-distressed) sample. This limits the generalizability of the IRAP toward distressed voice hearers and clinical samples in this domain. Therefore, future research should be carried non-clinical participants who score higher on distress permitting a functional analysis of the relationship between voice hearing and distress. Sixth, the use of rules within the IRAP studies has more recently been demonstrated to yield less reliable effects, therefore future studies should exclude rules from the parameters of the procedure and implement a shaping procedure when instructing participants (see Finn et al., 2016).

Future Directions. Given the current preliminary findings, our hope is to explore more complex aspects of the voice hearing experience that pertains to suffering, rather than merely its presence. This research could compare voices in distressed with non-distressed individuals, with a view to obtaining a greater understanding of the problematic relations involved in voice-related distress. Furthermore, it would be interesting to examine if a therapeutic intervention (such as Acceptance and Commitment Therapy for psychosis, see Bach & Hayes, 2002) could change these IRAP effects. Specifically, we would argue that existing interventions could be enhanced by identifying the problematic relations involved in voice hearing through functional analyses using the IRAP, and addressing these relations in interventions. We would also like to investigate whether the IRAP can predict the presence of voice hearing, voice-related distress, and clinical outcomes. These types of studies would provide a strong springboard from which an empirically-based functional-analytic account of voice hearing and psychosis could be built.

Compliance with Ethical Standards

Funding. This study was funded by FWO (Odysseus Type I Grant awarded to Prof. Dermot Barnes-Holmes).

Conflict of Interest. All authors declare that they have no conflict of interest.

Ethical approval. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent. Informed consent was obtained from all individual participants included in the study.

References

Bach, P., & Hayes, S. C. (2002). The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized controlled trial. *Journal of consulting and clinical psychology*, *70*(5), 1129-1139. doi:10.1037/0022-006X.70.5.1129.

Barnes-Holmes, D., Barnes-Holmes, Y., Hussey, I., & Luciano, C. (2016). Relational frame theory: Finding its historical and philosophical roots and reflecting upon its future development: An introduction to part II. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds), *The Wiley handbook of contextual behavioral science* (pp. 117-128), West Sussex, UK: Wiley-Blackwell.

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the implicit relational assessment procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist, 32*(7), 169-177.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *Psychological Record, 60*(3), 527–542.

Beavan, V., Read, J., & Cartwright, C. (2011). The prevalence of voice-hearers in the general population: A literature review. *Journal of Mental Health, 20*(3), 281–292. doi:10.3109/09638237.2011.562262.

Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine, 30*(5), 1131-1139.

Bond, F., Hayes, S., Baer, R., Carpenter, K., Guenole, N., Orcutt, H., … Zettle, R. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*(4), 676–88. doi:10.1016/j.beth.2011.03.007.

Carpenter, K. M., Martinez, D., Vadhan, N. P., Barnes-Holmes, D., & Nunes, E. V. (2013). Measures of attentional bias and relational responding are associated with behavioral treatment outcome for cocaine dependence. *Amrerican Journal of Drug and Alcohol Abuse, 38*(2), 146-154. doi:10.3109/00952990.2011.643986.

Chadwick, P., & Birchwood, M. (1994). The omnipotence of voices. A cognitive approach to auditory hallucinations. *British Journal of Psychiatry, 164*(2), 190-201.

Chadwick, P., Lees, S., & Birchwood, M. (2000). The revised Beliefs About Voices Questionnaire (BAVQ-R). *The British Journal of Psychiatry, 177,* 229–232.

Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist, 59*(7), 614-625. doi:10.1037/0003-066X.59.7.614.

Dorahy, M. J., Shannon, C., Seagar, L., Corr, M., Stewart, K., Hanna, D., ... Middleton, W. (2009). Auditory hallucinations in dissociative identity disorder and schizophrenia with and without a childhood trauma history: Similarities and differences. *Journal of Nervous and Mental Disease, 3*(12), 892-898. doi:10.1097/NMD.0b013e3181c299ea.

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), 119–22. doi:10.1016/j.schres.2006.04.024.

Eaton, W. W., Romanoski, A., Anthony, J. C., & Nestadt, G. (1991). Screening for psychosis in the general population with a explicit interview. *The Journal of Nervous and Mental Disease,179,* 689-693.

Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the Implicit Relational Assessment Procedure: The impact of three types of introductory rules. *The Psychological Record, 66*(2), 309-321. doi:10.1007/s40732-016-0173-4.

Fischer, S. (2013). Assessing the test-retest reliability of the IRAP in the context of sexuality. (Unpublished bachelor thesis). National University of Ireland Maynooth, Ireland.

Gauntlett-Gilbert, J., & Kuipers, E. (2005). Visual hallucinations in psychiatric conditions: appraisals and their relationship to distress. *British Journal of Clinical Psychology, 44,* 77–87. doi:10.1348/014466504x19451.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Haddock, G., McCarron, J., Tarrier, N., & Farragher, E. B. (1999). Scales to measure dimensions of hallucinations and delusions: The Psychotic Symptom Rating Scales (PSYRATS). *Psychological Medicine, 29,* 879–889.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A post Skinnerian account of human language and cognition.* New York: Plenum Press.

Hussey, I., Barnes-Holmes, D., & Booth, R. (2016). Individuals with current suicidal ideation demonstrate implicit "fearlessness of death". *Journal of Behavior Therapy and Experimental Psychiatry, 51,* 1-9. doi:10.1016/j.jbtep.2015.11.003.

Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science, 4*(3), 157-162. doi:10.1016/j.jcbs.2015.05.001.

Jackson, L. J., Hayward, M., & Cooke, A. (2011). Developing positive relationships with voices: A preliminary Grounded Theory. *International Journal of Social Psychiatry, 57,* 487–495. doi: 10.1177/0020764010368624.

Johns, L. C., Kompus, K., Connell, M., Humpston, C., Lincoln, T. M., Longden, E., … Laroi, F. (2014). Auditory verbal hallucinations in persons with and without a need for care. *Schizophrenia Bulletin, 40*(4), S255-S264. doi:10.1093/schbul/sbu005.

Kim, S. H., Jung, H. Y., Hwang, S. S., Chang, J. S., Kim, Y., Ahn, Y. M., & Kim, Y. S. (2010). The usefulness of a self-report questionnaire measuring auditory verbal hallucinations. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 34*(6), 968-973. doi:10.1016/j.pnpbp.2010.05.005.

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

Mak, W. W. S., Poon, C. Y. M., Pun, L. Y. K., & Cheung, S. F. (2007). Meta-analysis of stigma and mental health. *Social Science & Medicine, 65*(2), 245-261. doi:10.1016/j.socscimed.2007.03.015.

Mawson, A., Berry, K., Murray, C., & Hayward, M. (2011). Voice hearing within the context of hearers' social worlds: An interpretative phenomenological analysis. *Psychology and Psychotherapy: Theory, Research and Practice, 84,* 256–272. doi:10.1348/147608310X524883.

McEnteggart, C., Barnes-Holmes, Y., & Adekuoroye (in press). The Effects of a Voice Hearing Simulation on Implicit Fear of Voices. *Journal of Contextual Behavioral Science.* doi:10.1016/j.jcbs.2016.06.003.

McEnteggart, C. Mellon, M., van Dongen, L., Barnes-Holmes, Y., Egger, J., Barnes-Holmes, D., Hendricks, A., & Harte, C. (under review). The emotional and behavioural responses to voices: An implicit approach. *Behavior Therapy.*

Miller, L. J., O'Connor, E., & DiPasquale, T. (1993). Patients' attitudes toward hallucinations. *American Journal of Psychiatry*, *150*, 584-584. doi:10.1176/ajp.150.4.584.

Morrison, A. P., Nothard, S., Bowe, S. E., & Wells, A. (2004). Interpretations of voices in patients with hallucinations and non-patient controls: A comparison and predictors of distress in patients. *Behaviour Research and Therapy, 42,* 1315-1323. doi:10.1016/j.brat.2003.08.009.

Nicholson, E., & Barnes-Holmes, D. (2012a). Developing an implicit measure of disgust propensity and disgust sensitivity: Examining the role of implicit disgust propensity and sensitivity in obsessive-compulsive tendencies. *Journal of Behavioural Therapy and Experimental Psychiatry, 43*(3), 922-930. doi:10.1016/j.jbtep.2012.02.001.

Nicholson, E., & Barnes-Holmes, D. (2012b). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record, 62*, 263-277. doi:10.1007/s40732-016-0176-1.

Nicholson, E., McCourt, A., & Barnes-Holmes, D. (2013). The implicit relational assessment procedure as a measure of obsessive beliefs in relation to disgust. *Journal of Contextual Behavioral Science, 2,* 23-30. doi:10.1016/j.jcbs.2013.02.002.

O'Brien, P., & Johns, L. (2013). A graded exposure intervention for distressing visual hallucinations in schizophrenia. *Behavioural and Cognitive Psychotherapy, 41,* 495-499. doi:10.1017/S1352465812001130.

Peters, E. R., Williams, S. L., Cooke, M. A., & Kuipers, E. (2012). It's not what you hear, it's the way you think about it: Appraisals as determinants of affect and behaviour in

voice hearers. *Psychological Medicine, 42*(7), 1507-1514. doi:10.1017/S0033291711002650.

Ratcliff, K., Farhall, J., & Shawyer, F. (2010). Auditory Hallucinations : A Review of Assessment Tools. *Clinical Psychology and Psychotherapy, 18*, 524-534. doi:10.1002/cpp.729.

Rössler, W., Riecher-Rössler, A., Angst, J., Murray, R., Gamma, A., Eich, D., ... Gross, V. A. (2007). Psychotic experiences in the general population: A twenty-year prospective community study. *Schizophrenia Research, 92*(1), 1-14. doi:10.1016/j.schres.2007.01.002.

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., & Day, R. (1986). Early manifestations and first contact incidence of schizophrenia in different cultures. *Psychological Medicine, 16*(4), 909-928.

Shawyer, F., Ratcliff, K., Mackinnon, A., Farhall, J., Hayes, S., & Copolov, D. (2007). The Voices Acceptance and Action Scale (VAAS): Pilot Data. *Journal of Clinical Psychology, 63*(6), 593–606. doi:10.1002/jclp.20366.

Slotema, C. W., Daalman, K., Blom, J. D., Diederen, K. M., Hoek, H. W., & Sommer, I. E. C. (2012). Auditory verbal hallucinations in patients with borderline personality disorder are similar to those in schizophrenia. *Psychological Medicine, 42*(9), 1873-1878. doi:10.1017/S0033291712000165.

Sommer, I. E., Daalman, K., Rietkerk, T., Diederen, K. M., Bakker, S., Wijkstra, J., & Boks, M. P. (2010). Healthy individuals with auditory verbal hallucinations: Who are they? 27 psychiatric assessments of a selected sample of 103 subjects. *Schizophrenia Bulletin, 36*(3), 633-641. doi:10.1093/schbul/sbn130.

Stefanis, N., Hanssen, M., Smirnis, N., Avramopoulos, D., Evdokimidis, I., Stefanis, C., … Van Os, J. (2002). Evidence that three dimensions of psychosis have a distribution in the general population. *Psychological medicine*, *32*(2), 347–58.

Tien, A. Y. (1991). Distribution of hallucinations in the population. *Social Psychiatry and Psychiatric Epidemiology, 26,* 287-292.

Vahey, N.A, Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry, 48,* 59-65. doi:10.1016/j.jbtep.2015.01.004.

van der Gaag, M., Hageman, M. C., & Birchwood, M. (2003). Evidence for a cognitive model of auditory hallucinations. *Journal of Nervous and Mental Disease, 191,* 542−545. doi: doi: 10.1016/j.cpr.2009.11.006.

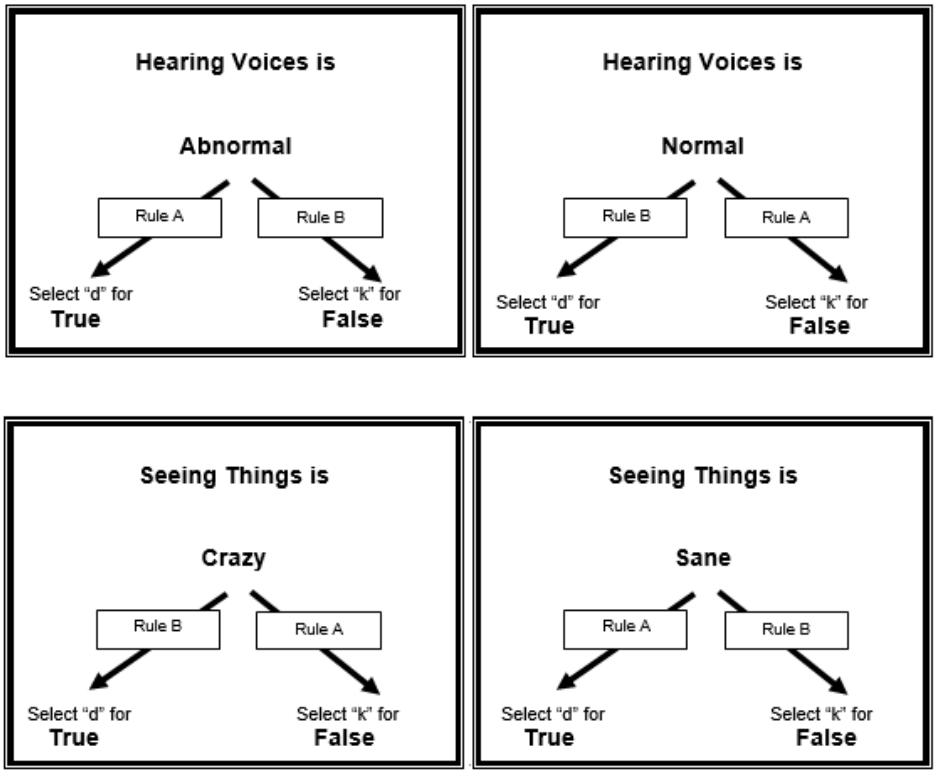
Hearing Voices is
Abnormal
Rule A
Rule B
Select "d" for
True
Select "k" for
False
Hearing Voices is
Normal
Rule B
Rule A
Select "d" for
True
Select "k" for
False
Seeing Things is
Crazy
Rule B
Rule A
Select "d" for
True
Select "k" for
False
Seeing Things is
Sane
Rule A
Rule B
Select "d" for
True
Select "k" for
False

Fig 1.

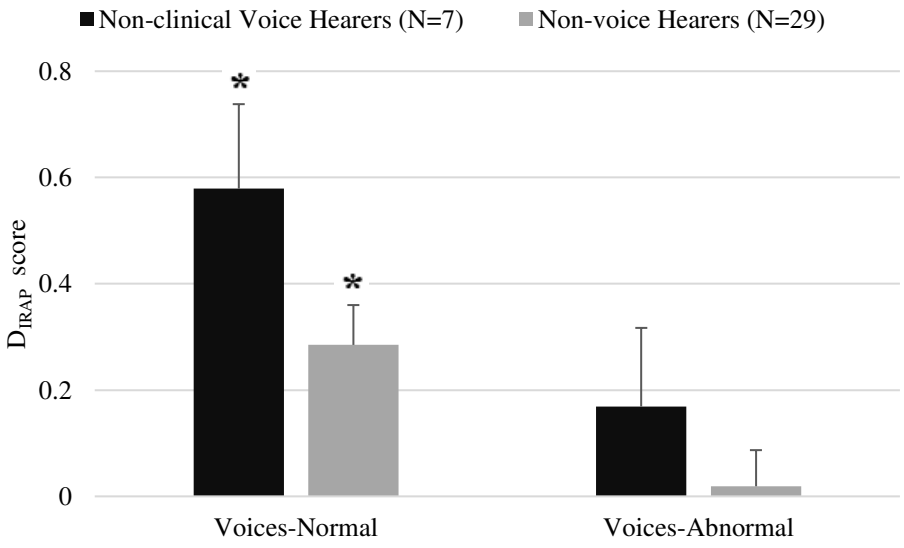
Non-clinical Voice Hearers (N=7)
Non-voice Hearers (N=29)
0.8
0.6
0.4
0.2
0
D IRAP score
*
*
Voices-Normal
Voices-Abnormal

Fig 2.

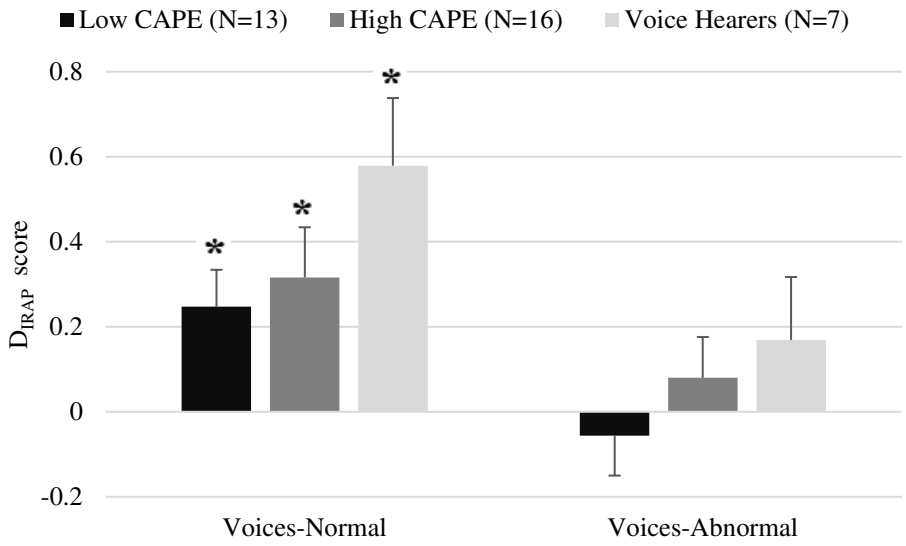
Low CAPE (N=13)
High CAPE (N=16)
Voice Hearers (N=7)
0.8
0.6
0.4
0.2
0
-0.2
D IRAP score
*
*
*
Voices-Normal
Voices-Abnormal

Fig 3.

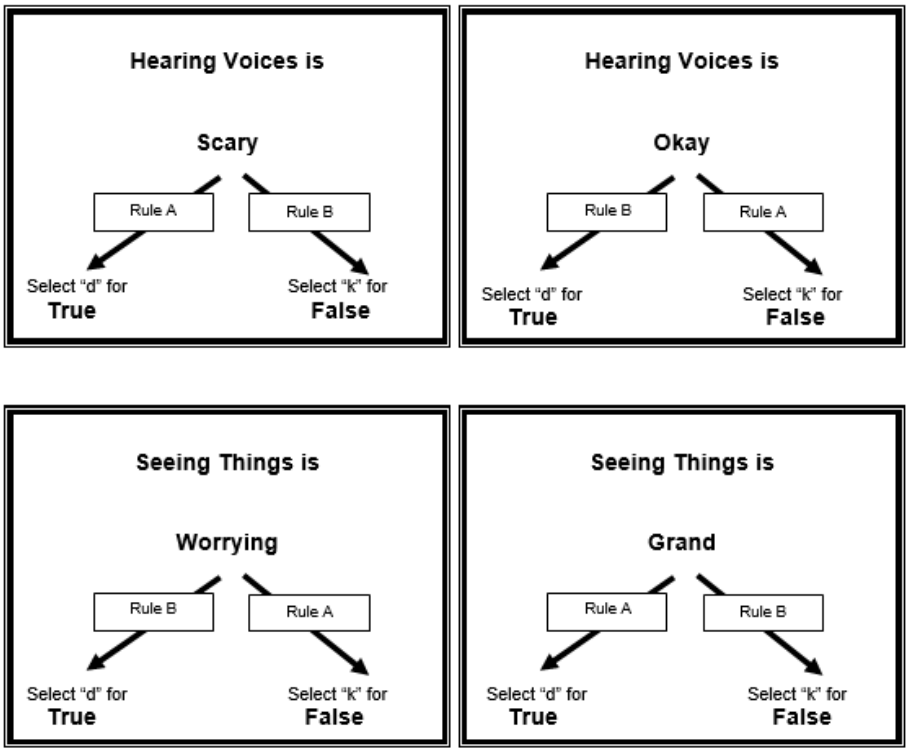
Hearing Voices is
Scary
Rule A
Rule B
Select "d" for
True
Select "k" for
False
Hearing Voices is
Okay
Rule B
Rule A
Select "d" for
True
Select "k" for
False
Seeing Things is
Worrying
Rule B
Rule A
Select "d" for
True
Select "k" for
False
Seeing Things is
Grand
Rule A
Rule B
Select "d" for
True
Select "k" for
False

Fig 4.

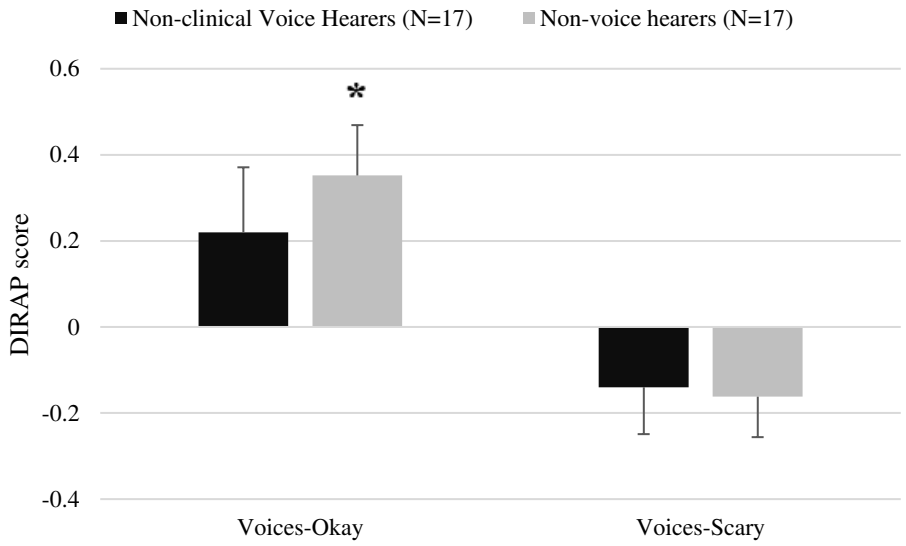
Non-clinical Voice Hearers (N=17)
Non-voice hearers (N=17)
0.6
0.4
0.2
0
-0.2
-0.4
DIRAP score
*
Voices-Okay
Voices-Scary

Fig 5.

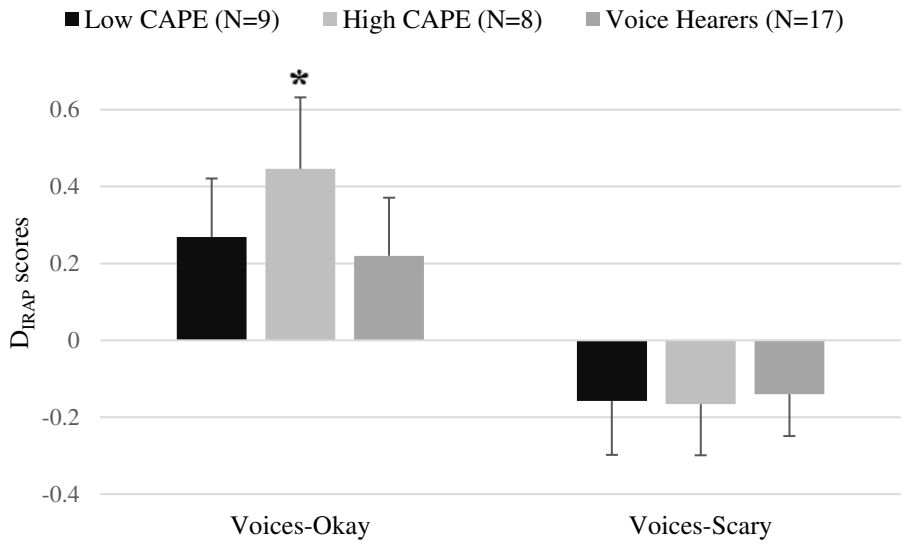
Low CAPE (N=9)
High CAPE (N=8)
Voice Hearers (N=17)
0.6
0.4
0.2
0
-0.2
-0.4
DIRAP scores
*
Voices-Okay
Voices-Scary

Fig 6.

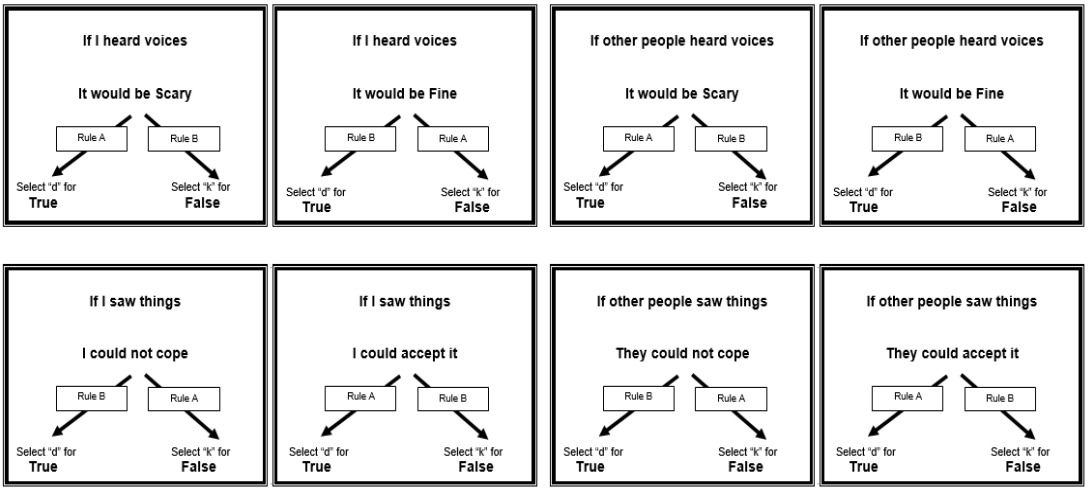
If I heard voices
It would be Scary
Rule A
Rule B
Select "d" for True
Select "k" for False
If I heard voices
It would be Fine
Rule B
Rule A
Select "d" for True
Select "k" for False
If other people heard voices
It would be Scary
Rule A
Rule B
Select "d" for True
Select "k" for False
If other people heard voices
It would be Fine
Rule B
Rule A
Select "d" for True
Select "k" for False
If I saw things
I could not cope
Rule B
Rule A
Select "d" for True
Select "k" for False
If I saw things
I could accept it
Rule A
Rule B
Select "d" for True
Select "k" for False
If other people saw things
They could not cope
Rule B
Rule A
Select "d" for True
Select "k" for False
If other people saw things
They could accept it
Rule A
Rule B
Select "d" for True
Select "k" for False

Fig 7.

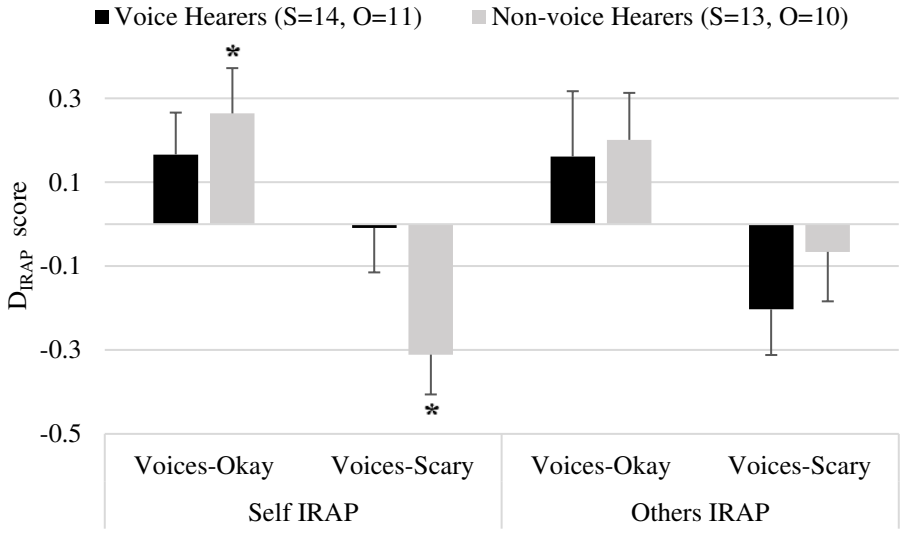
Voice Hearers (S=14, O=11)
Non-voice Hearers (S=13, O=10)
D IRAP score
0.3
0.1
-0.1
-0.3
-0.5
*
*
Voices-Okay
Voices-Scary
Voices-Okay
Voices-Scary
Self IRAP
Others IRAP

Fig 8.

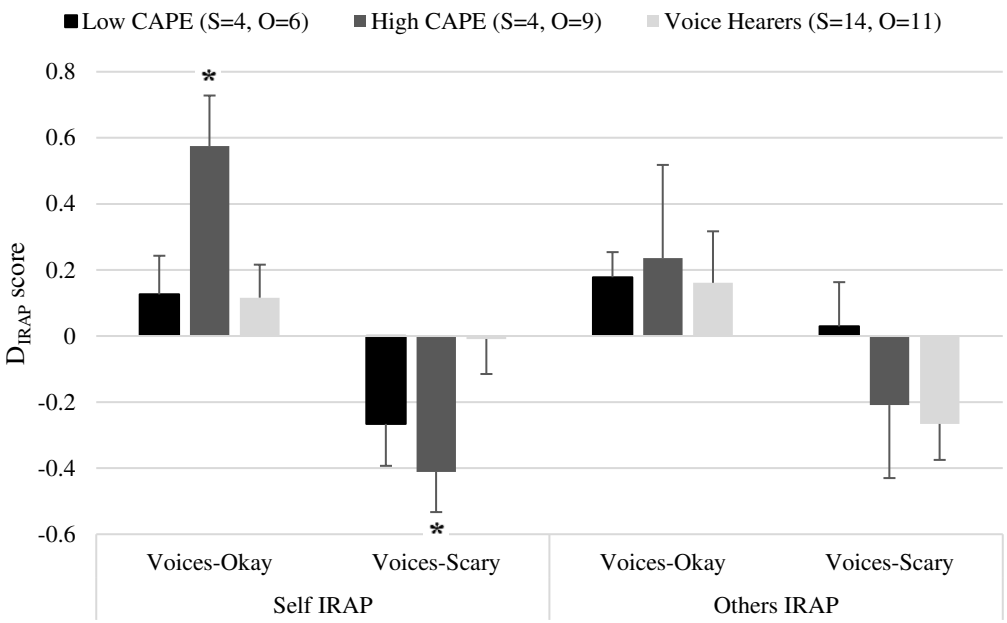
Low CAPE (S=4, O=6)
High CAPE (S=4, O=9)
Voice Hearers (S=14, O=11)
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
D IRAP score
*
*
Voices-Okay
Voices-Scary
Voices-Okay
Voices-Scary
Self IRAP
Others IRAP

Fig 9.

Table 1

Stimuli and Response Options of the Normality IRAP employed in Study 1

Normality IRAP	
Hearing Voices is	Seeing Things is
Normal Sane Acceptable	Abnormal Insane Crazy
True	False

Table 2

Explicit Data Summary for Study 1

Scales	**Voice Hearers (*N*=7)**	**Controls (*N*=29)**
	Mean (SD)	***Mean (SD)***
AAQ-II	24.57 (10.01)	22.66 (8.97)
CAPE		
CAPE positive dimension	2.45 (0.48)	2.0 (0.61)
CAPE depressive dimension	3.80 (1.58)	3.47 (1.18)
CAPE negative dimension	3.58 (0.91)	3.13 (0.94)
BAVQ-R		
Malevolence	7.29 (4.42)	-
Benevolence	10.29 (6.75)	-
Omnipotence	10.00 (7.66)	-
Emotional Resistance	6.43 (4.35)	-
Behavioural Resistance	13.28 (8.01)	-
Emotional Engagement	7.29 (6.02)	-
Behavioural Engagement	6.43 (4.11)	-
VAAS	47.71 (5.91)	-
AHRS	9.57 (8.02)	-

*Note. Maximum scores are: AAQ-II = 70; CAPE dimensions = 6; BAVQ-R: Mal, Ben & Omni = 18; Beh. Res. = 15, Emo. Res., Beh. Eng., Emo. Eng. = 12; VAAS = 155; AHRS = 44. Missing values for explicit measures which were not administered to the control participants are denoted by "-".Significance indicated by *(p<0.05).

Table 3

Stimuli and Response Options of the Fear IRAP employed in Study 2

Fear IRAP	
Hearing Voices is	Seeing Things is

Okay Fine Grand	Scary Distressing Worrying
True	False

Table 4

Explicit Data Summary for Study 2

Scales	**Voice Hearers (*N*=17)**	**Controls (*N*=17)**
	Mean (SD)	***Mean (SD)***
AAQ-II	22.77 (7.48)	19.65 (8.37)
CAPE		
CAPE positive dimension	2.23 (0.84)	2.05 (0.58)
CAPE depressive dimension	3.13 (1.29)	3.27 (1.21)
CAPE negative dimension	3.26 (1.56)	2.57 (0.76)
BAVQ-R		
Malevolence	6.71 (5.03)	-
Benevolence	11.12 (8.02)	-
Omnipotence	8.00 (5.71)	-
Emotional Resistance	6.24 (5.54)	-
Behavioural Resistance	10.00 (6.71)	-
Emotional Engagement	7.41 (5.10)	-
Behavioural Engagement	5.88 (4.21)	-
VAAS	47.65 (15.72)	-
AHRS	11.12 (8.05)	-
DASS	28.94 (17.79)	29.18 (21.60)
Depression	6.82 (6.25)	7.88 (8.90)
Anxiety	7.29 (7.61)	8.24 (8.09)
Stress	15.18 (8.22)	13.41 (9.82)

*Note. Maximum scores are: AAQ-II = 70; CAPE dimensions = 6; BAVQ-R: Mal, Ben & Omni = 18; Beh. Res. = 15, Emo. Res., Beh. Eng., Emo. Eng. = 12; VAAS = 155; AHRS = 44; DASS total = 126; DASS subscales = 42. Missing values for explicit measures which were not administered to the control participants are denoted by "-". Significant effects denoted by * ($p < 0.05$).

Table 5

Stimuli and Response Options of the two IRAPs employed in Study 3

Self IRAP		Others IRAP	
If I heard voices	If I saw things	If Other People heard voices	If Other People saw things
It would be fine I could accept it I could cope	It would be scary I could not accept it I could not cope	It would be fine They could accept it They could cope	It would be scary They could not accept it They could not cope
True	False	True	False

Table 6

Explicit Data Summary for Study 3

	Self IRAP		**Others IRAP**	
Scales	**Voice Hearers (*N*=14)**	**Controls (*N*=13)**	**Voice Hearers (*N*=11)**	**Controls (*N*=10)**
	Mean (SD)	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***
AAQ-II	20.86 (6.68)	22.00 (10.95)	27.00 (9.89)	21.70 (8.46)
CAPE				
CAPE positive dimension	3.04 (1.60)*	1.82 (0.57)*	2.47 (0.72)*	1.91 (0.47)*
CAPE depressive dimension	1.99 (1.13)	1.81 (1.10)	1.84 (1.43)*	5.68 (1.04)*
CAPE negative dimension	3.14 (1.47)	3.32 (1.15)	2.22 (0.73)	1.94 (0.62)
BAVQ-R				
Malevolence	9.57 (4.72)	-	7.09 (4.25)	-
Benevolence	13.79 (6.66)	-	12.55 (6.93)	-
Omnipotence	10.93 (4.88)	-	10.09 (4.66)	-
Emotional Resistance	8.21 (3.77)	-	6.46 (3.75)	-
Behavioural Resistance	10.50 (5.86)	-	10.55 (6.66)	-
Emotional Engagement	9.93 (6.03)	-	9.91 (6.49)	-
Behavioural Engagement	7.21 (3.40)	-	5.73 (4.41)	-
VAAS	58.07 (20.06)	-	48.00 (7.28)	-
AHRS	10.43 (8.67)	-	14.55 (6.88)	-

*Note. Maximum scores are: AAQ-II = 70; CAPE dimensions = 6; BAVQ-R: Mal, Ben & Omni = 18; Beh. Res. = 15, Emo. Res., Beh. Eng., Emo. Eng. = 12; VAAS = 155; AHRS = 44. Missing values for explicit measures which were not administered to the control participants are denoted by "-". Significant differences denoted by * (p<0.05).