Journal of Psychiatric Research 89 (2017) 48–54

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/psychires

Improving measurement of attributional style in schizophrenia; A psychometric evaluation of the Ambiguous Intentions Hostility Questionnaire (AIHQ)

Benjamin Buck [a, *], Colin Iwanski [a], Kristin M. Healey [a], Michael F. Green [c, d], William P. Horan [c, d], Robert S. Kern [c, d], Junghee Lee [c, d], Stephen R. Marder [c, d], Steve P. Reise [e], David L. Penn [a, b]

[a] University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[b] Australian Catholic University, Melbourne, Victoria, Australia
[c] Semel Institute for Neuroscience and Human Behavior University of California, Los Angeles, USA
[d] VA Greater Los Angeles Healthcare System, Los Angeles, VA, USA
[e] Department of Psychology, University of California, Los Angeles, USA

ARTICLE INFO

Article history:
Received 29 August 2016
Received in revised form
29 December 2016
Accepted 6 January 2017

Keywords:
Schizophrenia
Social cognition
Functioning
Psychometrics

ABSTRACT

While attributional style is regarded as a core domain of social cognition, questions persist about the psychometric characteristics of measures used to assess it. One widely used assessment of attributional style is the Ambiguous Intentions Hostility Questionnaire (AIHQ). Two limitations of the AIHQ include (1) a possible restricted range resulting from too few and too homogenous item scenarios, and (2) use of rater scores that are cumbersome and time-consuming to score and have unknown incremental validity. The present study evaluated the psychometric properties of the AIHQ while concurrently testing changes aiming to improve the scale, in particular expansion of the number of self-report items and removal of the rater-scored items. One hundred sixty individuals diagnosed with schizophrenia and 58 healthy controls completed the full AIHQ along with measures of symptoms, functioning, and verbal intelligence. The AIHQ – particularly the self-reported blame score – demonstrated adequate internal consistency, test-retest reliability, and distinguished patients from controls. It also was significantly related to clinically-rated hostility and suspiciousness symptoms, and correlated with functional capacity even after controlling for verbal intelligence. Incremental validity analyses suggested that a higher number of self-report items strengthens relationships to outcomes in a manner that justifies this expansion, while rater-scored items had mixed results in providing additional information beyond self-report in the AIHQ.

1. Introduction

Individuals with schizophrenia are consistently impaired in social cognition (Savla et al., 2013), which is defined as "the mental operations that underlie social interactions, including perceiving, interpreting, and generating responses to the intentions, dispositions, and behaviors of others" (Green et al., 2008, p. 1211). Social cognition is separable from general neurocognition (Allen et al., 2007; Van Hooren et al., 2008), a strong and consistent predictor of functioning (Couture et al., 2006; Fett et al., 2011), and has been shown to be responsive to psychosocial interventions (Kurtz and Richardson, 2012). Recent work has supported separation of two (Buck et al., 2016) or three (Mancuso et al., 2010) factors of social cognition in schizophrenia. One factor in each of these analyses has been attributional style or attributional bias, defined as the way in which individuals explain the causes, or make sense, of social events or interactions (Pinkham et al., 2013). Aberrant attributional style in psychosis has relates to positive (Combs et al., 2009) and hostility symptoms (An et al., 2010; Mancuso et al., 2010), as well as paranoia (Combs et al., 2009), depression (Mancuso et al., 2010), social conflict (Buck et al., 2016), and violence (Waldheter et al., 2005).

While attributional style is a key part of social cognition in

* Corresponding author. University of North Carolina at Chapel Hill, Davie Hall, CB #3270, Chapel Hill, NC 27599, USA.
E-mail address: bebuck@email.unc.edu (B. Buck).

http://dx.doi.org/10.1016/j.jpsychires.2017.01.004
0022-3956/

schizophrenia, it has proven difficult to measure. The Social Cognition Psychometric Evaluation study (SCOPE; Pinkham et al., 2013; 2016) identified attributional style as a core domain in the researcher survey and expert panel phase, but found limitations with the one measure of this domain that was selected for psychometric evaluation, the Ambiguous Intentions Hostility Questionnaire (AIHQ; Combs et al., 2007). The AIHQ presents participants with written vignettes depicting others' negative social acts without clear motives. These vignettes vary according to the apparent clarity of the other's motives, particularly whether such actions are accidental, ambiguous, or intentional. In the majority of research on the AIHQ – and in SCOPE – a version consisting only of the five ambiguous items has been used. After each vignette, participants complete three self-report anchored questions about the actor's intention, the actor's blameworthiness, and the participant's own anger, and answer two open-ended questions about their interpretation of the actor's motive and how the participant would respond to the situation. The self-report scores are totaled (*blame score*) and the open-ended responses are scored by an independent rater according to how hostile the interpretation of the other's motives (*hostility bias*) and aggressiveness of the participant's hypothetical response (*aggression bias*). Thus, in response to each vignette there is one self-report total (*blame score*) and two rater-scored items (*hostility bias and aggression bias*). These items are usually totaled to subscales according to the kind of scenario that elicited the response, i.e. accidental, ambiguous, or intentional scenarios. Three key limitations emerged from the SCOPE study. First, neither the blame scale nor either rater-scored items of the AIHQ had relationships to functional outcomes. Second, the hostility and aggression biases demonstrated relatively weak test-retest reliability. Another associated challenge with these rater-scored items is that they require additional training, resources and time for raters to score and establish inter-rater reliability. For these reasons, the AIHQ was not included as a measure for the final SCOPE battery (Pinkham et al., 2016).

However, given the importance of this construct, it is important to examine whether the AIHQ can be modified to address existing limitations. The present analysis tests two potential modifications. First, the lack of relationships to functional outcomes in SCOPE may be related in part to limited range. In the present study we examined the added value of including accidental – and not only ambiguous – scenarios (Combs et al., 2007). Second, the substandard test-retest reliability of the rater-scored items in SCOPE could be related to additional error from non-perfect rater agreement. Rater-scored items also add experimenter burden because they require additional training and time for ratings (Buck et al., 2016; Combs et al., 2009; Mancuso et al., 2010). Given these considerations, it is important to know whether these items enhance the scale beyond the existing self-report items. Overall, the present study provides a psychometric analysis of the AIHQ and potential modifications to the scale, particularly: (1) additional situational vignettes (i.e. including accidental as well as ambiguous vignettes) and (2) rater-scored items (i.e. in addition to blame score).

2. Methods

2.1. Participants

As a part of the Social Cognition and Functioning in Schizophrenia (SCAF) study (Green et al., 2013; Kern et al., 2013; Olbert et al., 2013), a subset of 160 individuals meeting DSM IV-TR criteria for schizophrenia and 58 community controls completed the full version of the AIHQ, as well as all other study measures. Both versions of the AIHQ (ambiguous and accidental vignettes) were administered at follow-up visits only to a subset of participants in the schizophrenia group (n = 89) and this group was used to determine test-retest reliability. Full demographic information for each group is listed in Table 1, and full recruitment procedures are reported elsewhere (Kern et al., 2013). The groups were significantly different in years of personal education ($t = -6.22$, $p < 0.001$) and marital status, X^2 (4, $N = 335$) = 38.77, $p < 0.001$. No other demographic variable significantly differed between groups.

2.2. Measures

2.2.1. Ambiguous Intentions Hostility Questionnaire (AIHQ; Combs et al., 2007)

The Ambiguous Intentions Hostility Questionnaire (AIHQ; Combs et al., 2007) consists of second-person vignettes of negative social situations with unknown cause (e.g., “you are walking by a group of young people who laugh as you pass by”). For each vignette, participants rate the following on Likert scales: the intentionality of the other's action (1–6), how angry it would make the participant feel (1–5), and how much he or she would blame the other (1–5). These three items are totaled for an overall blame score. Responses to each item are summed for each scenario; thus, total scores range from 3 to 16 with higher scores indicating greater blame, perceived intention, and anger. Additionally, participants provide two open-ended responses: an explanation of why the event occurred, and what they would do in response to the event. These items are evaluated by trained raters (on a 1 to 5 scale) according to the extent to which the participant interpreted the situation in a hostile manner (hostility bias; Combs et al., 2007) and the extent to which the individual reports aggression in his or her behavioral response (aggression bias; Combs et al., 2007).

The AIHQ was initially developed with 15 situations that varied in intentionality: five intentional scenarios (e.g. “While driving, the person in the car behind you honks their horn and then cuts you off.), five accidental scenarios (e.g. “A friend of yours slips on the ice knocking you onto the ground.”), and five ambiguous scenarios (e.g. “you walk past a bunch of teenagers at a mall and you hear them start to laugh.”). The present study examines the use of accidental and ambiguous items above, whereas the scale typically only uses ambiguous items. The intentional items were not collected in the present study as they are thought to prototypically elicit hostile or intentional attributions, thus are less valuable in pinpointing the cognitive styles specific to paranoia. For each set (the 5 ambiguous and 5 accidental items) of the AIHQ, the self-report total ranges from 15 to 80, and each rater-scored item ranges from 5 to 25, with higher scores indicating an increased tendency to see others' actions as hostile and an increased tendency to hypothetically respond in an aggressive manner.

2.2.2. Verbal intelligence

Verbal intelligence was assessed with the WRAT, a 42-item task involving participants to read words at varying levels of language difficulty. The WRAT, though brief, is a highly significant predictor of both verbal and full-scale IQ (Wiens et al., 1993).

2.2.3. Symptoms

The expanded version of the Brief Psychiatric Rating Scale (BPRS-E; Lukoff et al., 1986; Ventura et al., 1994) is a 24-item semi-structured diagnostic tool used to assess the presence and severity of a number of psychiatric symptoms in a clinical population. The ratings cover the previous two weeks and are based on the patients' behavior and responses to the administrators' questions. The present study examined the “hostility/suspiciousness factor” which is computed by summing responses for hostility, suspiciousness, and uncooperativeness. Total scores on the BPRS-E hostility/

Table 1
Participant demographics and tests for differences between the schizophrenia and non-clinical control samples.

	Group		
	Schizophrenia ($n = 160$)	Control ($n = 58$)	Test for differences
Age	41.69 (12.29)	42.74 (10.90)	$t = 0.58$, $p = 0.57$
Education (years)	12.45 (1.73)	14.29 (2.00)	$\mathbf{t = -6.22}$, $\mathbf{p < 0.001}^{***}$
Gender			$\mathbf{X^2 = 2.49}$, $\mathbf{p = 0.12}$
Male	117 (73.1%)	36 (62.1%)	
Female	43 (26.9%)	22 (37.9%)	
Race			$X^2 = 3.68$, $p = 0.45$
White 74 (46.3%) 35 (60.3%)			
Black 72 (45.0%) 19 (32.8%)			
Asian 7 (4.4%) 2 (3.4%)			
Multi-Racial 7 (4.4%) 2 (3.4%)			
Marital Status			$\mathbf{X^2 = 38.77}$, $\mathbf{p < 0.001}^{***}$
Married	9 (5.6%)	21 (36.2%)	
Single/Never Married	122 (76.2%)	19 (32.8%)	
Divorced	25 (15.6%)	14 (24.1%)	
Widowed	1 (0.6%)	1 (1.7%)	
Separated	3 (1.9%)	3 (5.2%)	
Previous hospitalizations	7.81 (10.19)	–	–
Age at first hospitalization	23.83 (7.69)	–	–
AIHQ self-report			
Ambiguous items	42.66 (12.45)	33.84 (10.48)	$t = 4.81$, $p < 0.001^{***}$
Accidental items	33.38 (11.54)	27.31 (9.41)	$\mathbf{t = 3.95}$, $\mathbf{p < 0.001}^{***}$
Total	76.04 (21.38)	61.16 (18.01)	$\mathbf{t = 5.12}$, $\mathbf{p < 0.001}^{***}$
AIHQ hostility bias			
Ambiguous items	10.71 (3.31)	8.82 (2.42)	$t = 4.57$, $p < 0.001^{***}$
Accidental items	7.48 (2.57)	6.96 (1.68)	$\mathbf{t = 1.74}$, $\mathbf{p = 00.08}$
Total	18.11 (4.66)	15.78 (3.14)	$\mathbf{t = 4.20}$, $\mathbf{p < 0.001}^{***}$
AIHQ aggression bias			
Ambiguous items	9.43 (1.99)	9.42 (1.35)	$t = 0.02$, $p = 0.99$
Accidental items	9.48 (3.14)	9.49 (2.79)	$\mathbf{t = -0.02}$, $\mathbf{p = 00.99}$
Total	18.92 (4.23)	18.91 (3.43)	$\mathbf{t = 0.00}$, $\mathbf{p = 00.99}$
Symptoms			
BPRS Positive	2.42 (1.01)	–	–
BPRS Depression	2.16 (1.04)	–	–
BPRS Negative	1.92 (0.94)	–	–
BPRS Agitation	1.52 (0.61)	–	–
BPRS Hostility/Suspiciousness	6.13 (2.59)	–	–
BPRS Total	48.23 (14.17)	–	–
Functional outcome			
MASC	3.53 (0.43)	–	–
RFS	16.66 (4.85)	24.83 (3.23)	$\mathbf{t = -14.25}$, $\mathbf{p < 0.001}^{***}$
UPSA-B	0.72 (0.14)	–	–

Note: Hospitalization data beyond those collected related to inclusion and exclusion criteria were not collected in the control group. In the control sample, sample size for analyses including the accidental items, $n = 57$. In the schizophrenia sample, sample sizes are consistent with the following: AIHQ ($n = 160$), RFS ($n = 158$), UPSA ($n = 155$), MASC ($n = 150$), BPRS ($n = 159$).
BPRS = The Brief Psychiatric Rating Scale Expanded; UPSA = UCSD Performance-based Skills Assessment, RFS = Role Functioning Scale, MASC = Maryland Assessment of Social Competence.

suspiciousness factor range from 3 to 21, with higher scores indicating greater hostility and suspiciousness.

2.2.4. Role functioning

The Role Functioning Scale (RFS; McPheeters, 1984) is an interview-rated measure of community functioning in the past month in four areas: (a) working productivity, (b) independent living, (c) family relationships, and (d) social network. Total scores on the RFS range from 4 to 28, with higher scores indicating better community role functioning.

2.2.5. Functional capacity

The Maryland Assessment of Social Competence (MASC; Bellack et al., 1994) is a performance-rated role-play used to assess participants' ability to solve common everyday problems in an interpersonal context. It consists of four 3-min conversations between the participant and a research staff member who role plays a person in the community. The four scenarios consist of the following: one initiating conversation with a casual acquaintance, two involving negotiation and compromise (e.g., asking for a second chance on a job), and one standing up for one's rights (e.g., talking to a landlord about a leaky roof). The interactions are videotaped for later scoring. Each scenario was coded on three dimensions using a 5-point Likert scale: verbal skill (speech content), non-verbal skill (such as eye contact and gestures), and overall effectiveness (the ability to maintain focus and achieve the goal of the scenario). Higher scores on the MASC indicate better performance, and are reported here by the average score across each Likert scale.

The UCSD Performance-based Skills Assessment (UPSA; Patterson et al., 2001) is a measure of functional capacity and assesses five skills necessary for community functioning: general organization, finance, social/communications, transportation, and household chores. The UPSA involves role-play tasks that are simulations of events that participants may encounter in the community. For example, in one task, participants are asked to demonstrate how they would leave a voicemail message for their doctor asking to reschedule their medical appointment for the following day. Higher scores on the UPSA indicate better

performance across the five general skill areas.

2.3. Procedure

Advanced graduate students and staff with experience working with individuals with schizophrenia conducted all assessments with participants at the two sites involved with the SCAF study: the University of North Carolina at Chapel Hill and the University of California Los Angeles via the VA Greater Los Angeles Healthcare System. A subset of twenty-five participants with schizophrenia (n = 25) and twenty five controls (n = 25) was randomly selected to test inter-rater reliability for the rater-scored items. For all subscales (hostility bias in ambiguous items, hostility bias in accidental items, aggression bias in ambiguous items and aggression bias in accidental items) intraclass coefficients were good (ICCs > 0.80).

2.4. Data analytic plan

The present study examined the following psychometric properties of the AIHQ: First, reliability was assessed in two ways: (1) examination of internal consistency through examination of Cronbach's alphas, and (2) test-retest reliability through Pearson's correlations for each AIHQ subscale at baseline and follow-up. Second, group differences were each examined in a 2 × 2 MANOVA examining blame score, hostility bias, and aggression bias as a function both of group and type of scenario (i.e. accidental or ambiguous). It was hypothesized that participants with schizophrenia would demonstrate an elevated hostile attribution bias on all subscales; no hypothesis was made for an interaction between group and scenario type. Third, convergent and discriminant validity were examined as Pearson correlations between AIHQ totals and the Hostility/Suspiciousness factor of the BPRS (convergent validity, significant relationship hypothesized) as well as the WRAT (discriminant validity analysis, no significant relationship hypothesized). Fourth, in order to examine external validity, Pearson correlations between AIHQ totals and all outcomes – functional capacity (MASC, UPSA-B) and role functioning (RFS) – were examined and hypothesized to be significant. In addition to these Pearson correlations, consistent with previous findings (Kalin et al., 2015), we examined whether the AIHQ might contribute to predictions of role functioning from assessments of social competence. Finally, we examined value of specific versions of the scale in predicting convergent and external validity outcomes, including (1) the value of the accidental self-report items above and beyond the ambiguous self-report items (i.e. the ten-scenario item AIHQ vs. the five-scenario item AIHQ) and (2) the added value of the rater-scored items beyond the self-report item totals (in the ten-scenario item version of the AIHQ). These analyses are conducted as hierarchical linear regressions of each added scale to all external and convergent validity outcomes controlling for the shorter version of the scale in each analysis.

3. Results

3.1. Reliability

With regard to internal consistency (Table 2), in both patient and control group samples, the blame score total had good to excellent internal consistency in both ambiguous and accidental scenarios, as well as combined across item scenario type. The rater-scored items, however, demonstrated reduced internal consistency, as the hostility and aggression biases had poor to fair internal consistency in both groups in both item scenario type conditions and combined total. These values were higher in the patient group than the control group, but still did not achieve psychometric standards for appropriate internal consistency for research tools (Nunnally, 1967; Pinkham et al., 2016). Following Pinkham and colleagues' psychometric evaluation in SCOPE, defining good test-retest reliability as falling within the range of 0.6–0.8, the blame score total demonstrated good test-retest reliability in all item scenarios. The rater-scored items, however, did not reach this standard (with the exception of the hostility bias in accidental scenarios). When combining the ambiguous and accidental scenario items, all test-retest correlations as well as internal consistency coefficients increased, as all test-retest reliability estimates ranged from good to excellent. Test-retest analyses could not be conducted on the control group as these individuals did not attend a follow-up visit.

Table 2
AIHQ scale reliability for the blame score (self-report) and hostility and aggression biases (rater-scored) in both patient and control groups with regard to test-retest reliability, inter-rater reliability, and internal consistency.

	Reliability	
	Test-retest reliability (r)	Interna consistencyl (α)
Accidental scenarios		
Patients		
Blame score	**0.67*****	0.85
Hostility bias	**0.70*****	**0.64**
Aggression bias	**0.53*****	**0.60**
Controls		
Blame score	-	0.86
Hostility item	–	0.29
Aggression item	–	0.52
Ambiguous scenarios		
Patients		
Blame score	**0.76*****	0.86
Hostility bias	**0.59*****	**0.52**
Aggression bias	**0.53*****	**0.55**
Controls		
Blame score	-	0.87
Hostility bias	–	0.28
Aggression bias	–	0.25
Ambiguous + Accidental scenarios		
Patients		
Blame score	**0.82*****	0.90
Hostility bias	**0.71*****	**0.63**
Aggression bias	**0.60*****	**0.66**
Controls		
Blame score	-	0.92
Hostility bias	–	0.36
Aggression bias	–	0.54

Note: In the schizophrenia sample, sample sizes are consistent with the following: AIHQ accidental and ambiguous (n = 160), RFS (n = 158), UPSA (n = 155), MASC (n = 150), BPRS (n = 159). When controlling for the WRAT, the relationship between the AIHQ blame score for patients and the BPRS factor was reduced to non-significant. All other significant relationships remained so after controlling for WRAT.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.10$.

3.2. Group differences

A mixed-model MANOVA was conducted to examine differences in AIHQ scores between groups (schizophrenia or non-patient control), as well as item type (ambiguous items or accidental items). First, in the self-report items, there was a significant effect for item-type condition, [F (1, 216) = 99.25, $p < 0.001$, $\eta^2_p = 0.32$] such that ambiguous scenarios ($M = 40.32$, $SD = 12.56$) were rated significantly more hostile and blameworthy than accidental scenarios ($M = 31.77$, $SD = 11.32$). There was also a significant effect for group [$F(1, 216) = 22.36$, $p < 0.001$, $\eta^2_p = 0.09$], as participants with schizophrenia ($M = 76.04$, $SD = 21.38$) rated all scenarios as more

hostile and blameworthy than did controls ($M = 61.16$, $SD = 18.01$). The interaction (clinical group x item type total) was a non-significant trend, [F (1, 216) = 2.99, $p = 0.09$, $\eta^2_p = 0.01$]; differences between ambiguous and accidental item totals were numerically larger for participants with schizophrenia ($M_{difference} = 9.28$, $SD = 10.94$) compared to controls ($M_{difference} = 6.53$, $SD = 8.52$).

With regard to the rater-scored hostility bias item, there was a significant effect for scenario type [F (1, 214) = 108.29, $p < 0.001$, $\eta^2_p = 0.34$] such that ambiguous scenarios ($M = 10.71$, $SD = 3.31$) elicited more hostile attributions than did accidental scenarios ($M = 7.28$, $SD = 2.22$). There was also a significant effect for group, [F (1, 213) = 12.41, $p = 0.001$, $\eta^2_p = 0.06$], as schizophrenia group participant responses ($M = 18.11$, $SD = 4.66$) were rated as more hostile than non-patient controls ($M = 15.78$, $SD = 3.14$). The interaction of group x item type was also significant, [F (1, 213) = 8.44, $p = 0.004$, $\eta^2_p = 0.04$], indicating that differences between ambiguous and accidental item totals were larger among participants with schizophrenia ($M_{difference} = 3.31$, $SD = 3.39$) than controls ($M_{difference} = 1.86$, $SD = 2.74$).

With regard to the rater-scored aggression bias item, there was no significant effect for scenario type [F (1, 214) = 0.07, $p = 0.79$, $\eta^2_p = 0.00$] group, [F (1, 213) = 0.00, $p = 0.99$, $\eta^2_p = 0.00$], or group x item type interaction, [F (1, 213) = 0.00, $p = 0.97$, $\eta^2_p = 0.00$. Means are presented in Table 1.

3.3. Convergent and discriminant validity

Regarding convergent validity there was a significant positive correlation between the Hostility/Suspiciousness factor of the BPRS and the blame score in ambiguous scenarios (Table 3). There was no relationship of this factor with hostility bias or aggression bias in either scenario type total or the blame score in accidental scenarios. When combining total scores on each item across ambiguous and accidental scenarios, only the blame score was significantly related to the BPRS Hostility/Suspiciousness factor. With regard to discriminant validity, in patients, higher blame scores across both item scenario types, as well as elevations in hostility and aggression bias in accidental scenarios were associated – though modestly – with poorer performance on the WRAT.

3.4. External validity

Correlations were computed between the AIHQ and measures of functional capacity and role functioning. As shown in Table 3, for the blame score, a greater tendency to blame others in ambiguous situations was associated with poorer functional capacity on the UPSA, while blaming others in accidental situations was associated with poorer scores on both measures of functional capacity. Combining all items across scenario type, the blame score was associated with all outcome measures with the exception of role functioning. With regard to the hostility bias, an increased tendency to interpret others' actions as hostile in ambiguous scenarios was related to role functioning, while doing so in accidental scenarios or across all scenarios was related to both measures of functional capacity and role functioning. Generating hypothetical aggressive responses (aggression bias) in response to both ambiguous and accidental scenarios (and combined total) was related to one measure of role functioning. Given the results of discriminant validity analyses, we repeated the external validity analyses controlling for WRAT total score. The results were unchanged.

3.5. Incremental validity of the accidental subscale and rater-scored items

To evaluate whether the addition of the accidental scenario items added incremental validity to the self-report items of the AIHQ, we examined model improvement adding accidental subscale items beyond ambiguous items for convergent and external validity markers (Table 4). Using hierarchical linear regression, accidental blame score totals were entered at a step following ambiguous blame score totals. Indeed, accidental items significantly improved models predicting both measures of functional capacity, the MASC ($\Delta R^2 = 0.04$, $p = 0.016$) as well as the UPSA-B ($\Delta R^2 = 0.09$, $p < 0.001$).

Based on these results in support of expanding the scale to include accidental in addition to ambiguous scenarios, we followed a similar procedure to examine the added value of the rater-scored items, the hostility bias and aggression bias. While entering both of these values after first entering the blame score total improved the model predicting the RFS significantly ($\Delta R^2 = 0.06$, $p = 0.01$), the other three models were not significant.

3.6. Predicting functioning together with AIHQ

Finally, we sought to examine whether the full-length AIHQ improved predictions of real-world functioning above and beyond social competence (Table 5). At the first step, the MASC and

Table 3
Convergent, discriminant and external validity analyses, measured as Pearson correlations of AIHQ blame scores (self-report) and hostility and aggression biases (rater-scored) with measures of symptoms and functioning (in the schizophrenia sample only).

	Verbal Intelligence	Psychiatric Symptoms	Functional Outcome		
	WRAT Total	BPRS Host./Susp.	MASC Total	RFS Total	UPSA-B Total
Ambiguous scenarios					
Blame score	-0.17*	0.17**	-0.12	-0.04	-0.27**
Hostility bias	-0.07	0.09	-0.11	**-0.18***	-0.13
Aggression bias	-0.11	-0.08	-0.12	-0.02	**-0.21****
Accidental scenarios					
Blame score	-0.21**	0.12	-0.23**	-0.14^	-0.40***
Hostility bias	**-0.19***	0.02	**-0.23****	**-0.19***	**-0.33*****
Aggression bias	**-0.16***	0.04	-0.08	-0.12	**-0.24****
Ambiguous + Accidental scenarios					
Blame score	-0.21**	0.16*	-0.19*	-0.09	-0.37***
Hostility bias	-0.13	0.06–**0.21***	**-0.26****	**-0.26****	
Aggression bias	**-0.17***	-0.01	-0.11	-0.10	**-0.28*****

Note: In the schizophrenia sample, sample sizes are consistent with the following: AIHQ accidental and ambiguous ($n = 160$), RFS ($n = 158$), UPSA ($n = 155$), MASC ($n = 150$), BPRS ($n = 159$). When controlling for the WRAT, the relationship between the AIHQ blame score for patients and the BPRS factor was reduced to non-significant. All other significant relationships remained so after controlling for WRAT.
***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ^$p < 0.10$.

Table 4
Incremental validity analyses examining relationships of symptoms and functional outcome with (1) self-report items in accidental scenarios controlling for ambiguous scenarios, and (2) rater-scored items controlling for self-report items.

	Psychiatric Symptoms						Functional Outcome					
	BPRS Total			MASC Total			RFS Total			UPSA-B Total		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Accidental scenarios, controlling for ambiguous												
Step 1– Ambiguous scenarios												
Blame score, Ambiguous	**0.21**	**0.09**	**0.19***	−0.00	0.00	0.12	−0.01	0.03	−0.04	**−0.00**	**0.00**	**−0.27****
	(R^2 = 0.04, p = 0.02)*			(R^2 = 0.01, p = 0.15)			(R^2 = 0.00, p = 0.66)			**(R^2 = .07, p = 0.001)****		
Step 2– Accidental scenarios												
Blame score, Ambiguous	0.09	0.11	0.08	0.00	0.00	0.02	0.03	0.04	0.07	−0.00	0.00	−0.05
Blame score, Accidental	*0.23*	*0.12*	*0.19^*	***−0.01***	***0.00***	***−0.24****	*−0.07*	*0.04*	*−0.18^*	***−0.00***	***0.00***	***−0.37*****
	(ΔR^2 = 0.02, p = 0.05)^			**(ΔR^2 = .04, p = 0.016)***			(ΔR^2 = 0.02, p = 0.08)^			**(ΔR^2 = .09, p < 0.001)*****		
Rater-scored items, controlling for blame score self-report (combined ambiguous and accidental items)												
Step 1– Self-report items												
Blame score, total	**0.15**	**0.05**	**0.22****	**−0.00**	**0.00**	**−0.20****	−0.03	0.02	−0.12	**−0.00**	**0.00**	**−0.36*****
	(R^2 = .05, p = 0.005)**			**(R^2 = .04, p = 0.016)****			(R^2 = 0.02, p = 0.14)			**(R^2 = .13, p < 0.001)*****		
Step 2–Rater-scored items												
Blame score, total	**0.18**	**0.06**	**0.27****	−0.00	0.00	−0.11	0.01	0.02	0.04	**−0.00**	**0.00**	**−0.28****
Hostility Bias	*−0.09*	*0.28*	*−0.03*	*−0.01*	*0.01*	*−0.15*	***−0.28***	***−0.10***	***−0.27****	*−0.00*	*0.00*	*−0.06*
Aggression Bias	*−0.31*	*0.28*	*−0.09*	*−0.00*	*0.01*	*−0.04*	*−0.06*	*0.10*	*0.05*	*−0.00*	*0.00*	*−0.14*
	(ΔR^2 = 0.01, p = 0.51)			(ΔR^2 = 0.02, p = 0.28)			**(ΔR^2 = 0.06, p = 0.01)***			(ΔR^2 = 0.02, p = 0.18)		

Note: All of the above analyses examine only the schizophrenia sample; sample sizes are consistent with the following: AIHQ accidental and ambiguous (n = 160), RFS (n = 158), UPSA (n = 155), MASC (n = 150), BPRS (n = 159).
***p < 0.001, **p < 0.01, *p < 0.05, ^p < 0.10.

significantly predicted RFS total scores, $F(2, 143) = 16.57$, $p < 0.001$, $R^2 = 0.10$, the addition of the AIHQ totals (combined across scenario type, blame, hostility and aggression scale), did not improve the model, $\Delta R^2 = 0.06$, $p = 0.02$.

4. Discussion

Based on the results of this study, the AIHQ appears to be appropriate for use for individuals with schizophrenia, and further, its psychometric characteristics can be strengthened with measure modification. In particular, the present study supports the inclusion of blame score items related to both ambiguous and accidental scenarios, but provides mixed evidence for the continued inclusion of the rater-scored hostility and aggression biases, as the modest gains provided by these items may not justify additional time added by their administration and scoring. The AIHQ blame score showed good internal consistency among both participants with schizophrenia and controls, while the rater-scored items had lower internal consistency coefficients in both groups. Also, while test-retest reliability coefficients were sufficient for both blame score and rater-scored items, they were lower among hostility and aggression biases. These findings are consistent with previous concerns about the internal consistency and stability over time of rater-scored items of the AIHQ (Pinkham et al., 2016). With regard to group differences, in general, participants with schizophrenia demonstrated an elevated hostile attribution bias on the AIHQ compared to controls, and particularly in ambiguous, rather than accidental self-report and hostility bias items.

Table 5
Hierarchical linear regression predicting role functioning from social competence and AIHQ scale totals.

Variable	B	SE B	β
Step 1– Social competence			
MASC Total	***3.61***	***0.89***	***0.32*****
Step 2– AIHQ totals, all scales			
MASC Total	**3.08**	**0.89**	**0.27****
AIHQ Blame score, total	*−0.10*	*0.10*	*−0.09*
AIHQ Hostility bias, total	***−0.27***	***0.10***	***−0.26****
AIHQ Aggression bias, total	*0.02*	*0.02*	*0.08*

Note: Sample size for this analysis = 148. $R^2 = 0.10$ ($p < 0.001$) for Step 1, $\Delta R^2 = 0.06$ ($p = .02$). Total model $R^2 = 0.16$, $p < 0.001$.
^p < 0.10; *p < 0.05; **p < 0.01; ***p < 0.001.

Convergent validity analyses were consistent with previous studies (An et al., 2010; Buck et al., 2016; Combs et al., 2009; Mancuso et al., 2010) and showed that self-reported blame score is related to clinically-rated hostility and suspiciousness, while rater-scored hostility and aggression were not. Discriminant validity results revealed that the AIHQ was modestly related to a brief measure of premorbid verbal intelligence. On one hand, this appears a limitation for discriminant validity, which contradicts previous work (Mehta et al., 2014). On the other hand, social cognition and neurocognition have some shared variance (Fett et al., 2011; Pinkham et al., 2016). Finally, external validity analyses show a significant relationship of blame score and rater-scored items to functional capacity and role functioning, and these relationships persisted while controlling for verbal intelligence. The AIHQ also added variance to predictions of role functioning from measures of social competence, however, as has been shown with other skill-based domains of social cognition (Kalin et al., 2015). This suggests that – differs from the SCOPE results (Pinkham et al., 2016) – similar to other domains of social cognition (e.g. theory of mind, emotion processing), attributional style might predict functioning independent of neurocognition.

The present study specifically examines ways to improve the scale that might address previously identified limitations (Pinkham et al., 2016). In particular, incremental validity analyses suggest that extending the scale to include both ambiguous and accidental scenario items provides the scale with stronger relationships to two measures of functional capacity. This extension of the scale could be executed simply and without additional training. The rater-scored items present a more complicated question. On one hand, we found that rater-scored items are significantly but modestly related to community role functioning when controlling for self-report items; they also improve models predicting role functioning from assessments of social competence. On the other hand, these items are more cumbersome as they require additional time for both rating and training of raters. They also present with lower internal

consistency and test-retest reliability coefficients, are unrelated to hostility and suspiciousness symptoms, and are no more related to measures of functional capacity than the self-report alone.

This study has several limitations. First, order effects might have affected results, as accidental items were always read after the ambiguous items in the AIHQ. Second, the present study doesn't compare the value of additional ambiguous items to the value of additional accidental items, and it could be the case that additional items of any variety improve psychometric characteristics of the scale. Additional studies could examine additional kinds of items as well, including AIHQ intentional scenarios, which were not collected in the current study. Third, comparisons of convergent relationships between the schizophrenia group and controls weren't possible as the control group was not administered any measure of hostility or paranoia.

Overall, the present study suggests that the AIHQ is a psychometrically sound instrument of hostile attribution bias in schizophrenia, however, should be administered with more self-report items that vary across situational contexts. The expansion of self-report items presents a low-cost and minimally burdensome way to improve validity of the scale. Because the psychometric support for the rater scores is more mixed, recommendations with regard to these scales are more complicated. Rater-scored items can be used in circumstances in which more fine-grained hostile attribution measurement is desired by investigators and additional time and resources are available. The present study provides information to guide that choice.

Acknowledgements

The researchers would like to thank all of participants in this study and the following individuals for aiding with data collection and management: Mark McGee (UCLA), Charles Olbert (UNC), Betty Rupp (UNC), and Yusra Iftikhar (UNC). Funding for this project came from NIMH grants MH087618, MH43292, and MH065707 to Dr. Green.

References

Allen, D.N., Strauss, G.P., Donohue, B., van Kammen, D.P., 2007. Factor analytic support for social cognition as a separable cognitive domain in schizophrenia. Schizophr. Res. 93, 325–333.

An, S.K., Kang, J.I., Park, J.Y., Kim, K.R., Lee, S.Y., Lee, E., 2010. Attribution bias in ultra-high risk for psychosis and first-episode schizophrenia. Schizophr. Res. 118, 54–61.

Bellack, A.S., Sayers, M., Mueser, K.T., Bennett, M., 1994. Evaluation of social problem solving in schizophrenia. J. Abnorm. Psychol. 103, 371–378.

Buck, B., Healey, K.M., Gagen, E.C., Roberts, D.L., Penn, D.L., 2016. Social cognition in schizophrenia; Factor structure, clinical and functional correlates. J. Ment. Health 25, 330–337.

Combs, D., Penn, D., Wicher, M., Waldheter, E., 2007. The ambiguous intentions hostility questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. Cogn. Neuropsy 12, 128–143.

Combs, D., Penn, D., Michael, C., Basso, M., Wiedeman, R., Siebenmorgan, M., Tiegreen, J., Chapman, D., 2009. Perceptions of hostility by persons with and without persecutory delusions. Cogn. Neuropsy 14, 30–52.

Couture, S.M., Penn, D.L., Roberts, D.L., 2006. The functional significance of social cognition in schizophrenia: a review. Schizophr. Bull. 32, 44–63.

Fett, A.K.J., Viechtbauer, W., Penn, D.L., van Os, J., Krabbendam, L., 2011. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. Neurosci. Biobehav. Rev. 35, 573–588.

Green, M.F., Lee, J., Ochsner, K.N., 2013. Adapting social neuroscience measures for schizophrenia clinical trials, Part 1: ferrying paradigms across perilous waters. Schizophr. Bull. 39, 1192–1200.

Green, M.F., Penn, D.L., Bentall, R., Carpenter, W.T., Gaebel, W., Gur, R.C., Kring, A.M., Park, S., Silverstein, S.M., Heinssen, R., 2008. Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. Schizophr. Bull. 34, 1211–1220.

Kalin, M., Kaplan, S., Gould, F., Pinkham, A.E., Penn, D.L., Harvey, P.D., 2015. Social cognition, social competence, negative symptoms and social outcomes: Inter-relationships in people with schizophrenia. J. Psychiat. Res. 68, 254–260.

Kern, R.S., Penn, D.L., Lee, J., Horan, W.P., Reise, S.P., Ochsner, K.N., Marder, S.R., Green, M.F., 2013. Adapting social neuroscience measures for schizophrenia clinical trials, part 2: trolling the depths of psychometric properties. Schizophr. Bull. 39, 1201–1210.

Kurtz, M.M., Richardson, C.L., 2012. Social cognitive training for schizophrenia: a meta-analytic investigation of controlled research. Schizophr. Bull. 38, 1092–1104.

Lukoff, D., Nuechterlein, K.H., Ventura, J., 1986. Manual for the Expanded Brief Psychiatric Rating Scale.

Mancuso, F., Horan, W., Kern, R., Green, M., 2010. Social cognition in psychosis: multidimensional structure, clinical correlates, and relationship with functional outcome. Schizophr. Res. 125, 143–151.

Mehta, U.M., Bhagyavathi, H.D., Thirthalli, J., Kumar, K.J., Gangadhar, B.N., 2014. Neurocognitive predictors of social cognition in remitted schizophrenia. Psychiatr. Res. 219, 268–274.

McPheeters, H.L., 1984. Statewide mental health outcome evaluation: a perspective of two southern states. Comm. Ment. Health J. 20, 44–55.

Nunnally, J.C., Bernstein, I.H., Berge, J.M.T., 1967. Psychometric Theory, vol. 226. McGraw-Hill, New York.

Olbert, C.M., Penn, D.L., Kern, R.S., Lee, J., Horan, W.P., Reise, S.P., Ochsner, K.N., Marder, S.R., Green, M.F., 2013. Adapting social neuroscience measures for schizophrenia clinical trials, Part 3: fathoming external validity. Schizophr. Bull. 39, 1211–1218.

Patterson, T.L., Goldman, S., McKibbin, C.L., Hughs, T., Jeste, D.V., 2001. UCSD performance-based skills assessment: development of a new measure of everyday functioning for severely mentally ill adults. Schizophr. Bull. 27, 235–245.

Pinkham, A., Penn, D., Green, M., Buck, B., Healey, K., Harvey, P., 2013. The social cognition psychometric evaluation study: results of the expert survey and rand panel. Schizophr. Bull. 39, 1–11.

Pinkham, A.E., Penn, D.L., Green, M.F., Harvey, P.D., 2016. Social cognition psychometric evaluation (SCOPE): results of the initial psychometric study. Schizophr. Bull. 42, 494–504.

Savla, G.N., Vella, L., Armstrong, C.C., Penn, D.L., Twamley, E.W., 2013. Deficits in domains of social cognition in schizophrenia: a meta-analysis of the empirical evidence. Schizophr. Bull. 39, 979–992.

Van Hooren, S., Versmissen, D., Janssen, I., Myin-Germeys, I., à Campo, J., Mengelers, R., van Os, J., Krabbendam, L., 2008. Social cognition and neurocognition as independent domains in psychosis. Schizophr. Res. 103, 257–265.

Ventura, J., Green, M.F., Shaner, A., Liberman, R.P., 1994. Training and quality assurance with the brief psychiatric rating scale: “the drift busters.” Int. J. Meth. Psychol. Res. 3, 221–244.

Waldheter, E., Jones, N., Johnson, E., Penn, D., 2005. Utility of social cognition and insight in the prediction of inpatient violence among individuals with a severe mental illness. J. Nerv. Ment. Dis. 193, 609–618.

Wiens, A.N., Bryan, J.E., Crossen, J.R., 1993. Estimating WAIS-R FSIQ from the national adult reading test-revised in normal subjects. Clin. Neuropsyc 7, 70–84.