AUTOMATIC AND CONTROLLED COMPONENTS OF ATTRIBUTION BIASES IN SCHIZOPHRENIA; EXAMINATION OF NOVEL MEASURES OF INTENTIONALITY AND IMMORALITY BIAS

Benjamin Edwards Buck

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Ph.D. in the Clinical Psychology Program of the Department of Psychology and Neuroscience in the College of Arts and Sciences.

Chapel Hill
2017

Approved by:

Gabriel Dichter

Kurt Gray

Deborah Jones

B. Keith Payne

David L. Penn

ABSTRACT

Benjamin Edwards Buck: Automatic And Controlled Components Of Attribution Biases In Schizophrenia; Examination Of Novel Measures Of Intentionality And Immorality Bias
(Under the direction of David L. Penn)

Social cognition is a strong predictor of outcome in schizophrenia and it is responsive to psychosocial intervention. In schizophrenia research, it has been shown to comprise two categories of cognitive variables, skills and biases. Existing measures of one core social cognitive bias – hostile attribution bias – do not allow access to assessment of the processes underlying such judgments, they conflate multiple constructs, and relationships to criterion outcomes are modest. To address these limitations, the present study applied two innovative paradigms from social psychology – the intentionality bias (Rosset, 2008) and the immorality bias (Hester, Payne & Gray, 2017) – to a sample of individuals with schizophrenia or schizoaffective disorder and a comparison group of non-clinical controls. Results suggested that individuals with schizophrenia do present with an elevated bias to interpret intentionality in others' ambiguous actions, and this bias appears modestly related to interpersonal conflict, paranoia, hostile cognition and behavior, though unrelated to positive symptoms. The immorality bias, on the other hand, was not elevated in schizophrenia or paranoia, nor was it related to any disorder-related processes. In examining parameters separating controlled and automatic processes, the present study raises a number of questions about measurement of biases and dual-process procedures in schizophrenia.

To Addie, my family and my community, without whom none of this would have been possible.

ACKNOWLEDGEMENTS

This project was made possible by David Penn, Amy Pinkham, and Phil Harvey, the principal investigators of the SCOPE study. That study was funded by the National Institute of Mental Health (NIMH). I would like to thank all of the individuals who participated in this study and the following individuals for their assistance with data collection and management: Grant Hardaway (UTD), Katie Kemp (UTD), Lana Nye (UNC), Grace Lee Simmons (UNC), Sara Kaplan (UM), and Craig Winter (UM).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIHQ	Ambiguous Intentions Hostility Questionnaire
ANOVA	Analysis of Variance
BLERT	Bell Lysaker Emotion Recognition Task
CIST	Categorizing Immoral Sentences Task
CTRL	Control group participant
ER-40	Penn Emotion Recognition Task (40 Item)
IBT	Intentionality Bias Task
OSCARS	Observable Social Cognition: A Rating Scale
PADS	Persecution and Deservedness Scale
PANSS	Positive and Negative Syndrome Scale
PDP	Process Dissociation Procedure
PID-5	The Personality Inventory for DSM-5
SCOPE	Social Cognition Psychometric Evaluation Study
SCZ	Schizophrenia group participant
SLOF	Specific Levels of Functioning Scale
SSPA	Social Skills Performance Assessment
TASIT	The Awareness of Social Inferences Test
UPSA	UCSD Performance Skills Assessment
WRAT	Wide Range Achievement Test

CHAPTER 1: INTRODUCTION

Social Cognition in Schizophrenia

Schizophrenia, a severe mental disorder that affects nearly 1% of the world's population (Saha et al., 2005), is characterized by the positive symptoms of hallucinations, delusions, or disorganization, the negative symptoms of affect flattening, anhedonia, avolition or alogia, and significant impairment in vocational or social functioning (Bellack et al., 2007; Couture, Penn & Roberts, 2006). Early research work on the treatment of schizophrenia aimed at improvements in symptom management with psychotropic medications (Eckman et al., 1992; Falloon et al., 1982). While medications have been effective in managing positive symptoms, they have significant side effects (Leucht et al., 1999), and only provide modest improvements in functioning (e.g. work attendance and performance, social engagement, etc.) and negative symptoms (Sergi et al., 2007; Swartz et al., 2007).

Thus, recent developments in schizophrenia research have taken two aims: first, identify any measurable cognitive abilities that provide reliable predictions of functioning (Fett et al., 2011; Holthausen et al., 2007) and second, develop and evaluate evidence-based interventions that best remediate these domains (Gold, 2004; Kurtz & Richardson, 2012). The most well-established predictor of functioning in schizophrenia, neurocognition (Green et al., 2004; Twamley, Jeste & Bellack, 2003), has proven useful in generating treatments of the disorder (i.e. cognitive remediation; McGurk et al., 2007), but still only predicts a portion of the variance in functional outcomes (Velligan et al., 1997). More recently, other promising domains have been

identified to address this limitation, one of these being specifically *social* cognition (Penn et al., 1997).

Individuals with schizophrenia are consistently impaired in social cognition (Savla et al., 2013), which is defined in the clinical research on schizophrenia 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 statistically and theoretically from general neurocognition (Allen et al., 2007; van Hooren et al., 2008), is itself a robust predictor of concurrent (Couture, Penn & Roberts, 2006; Fett et al., 2011) and prospective (Horan et al., 2012) functioning, and is largely responsive to psychosocial interventions (Kurtz & Richardson, 2011). As a result of this, research into social cognition has been rapidly growing; the number of PsycINFO citations per year using "social cognition" and "schizophrenia" as key terms more than quintupled in the five years between 2002 and 2007 (Green & Leitman, 2008).

Differentiating social cognitive biases and social cognitive skills

Social cognition in schizophrenia comprises two categories: (1) abilities to correctly interpret social information, or *social cognition skills* and (2) specific patterns in open-ended interpretations of social situations, or *social cognitive biases* (Mancuso et al., 2011; Roberts & Pinkham, 2013). Each social cognition *skill* comprises a singular ability to arrive at a clear correct answer, thus each skill presents an area in which an individual is either impaired or skilled. These skills – which include abilities like emotion perception (Kohler et al., 2009), theory of mind (Bora, Yucel & Pantelis, 2009), and social perception (Sergi et al., 2006) – correlate highly with neurocognition (Fett et al., 2011; Sergi et al., 2007; van Hooren et al., 2008), independent living skills, social skills, and social functioning (Mancuso et al., 2011).

These social cognition skills are either unifactorial (Browne et al., 2016; Buck et al., 2016a), or divide across two-factors according to level of the complexity of judgment involved (i.e. low-level cue detection and higher order processes; Mancuso et al., 2011).

Differently, *social cognition biases* do not assess one's ability to *correctly* respond in a right-or wrong determination, but rather examine the *style* with which one tends to respond in certain social circumstances. In this way, pathological responding is not identified solely through so-called "deficits," but instead, generally extreme response patterns. Five social cognitive biases have been identified in the schizophrenia literature: (1) the bias to attribute negative events to the external world and positive events to oneself, or *externalizing bias* (EB; Bentall, Kinderman & Kaney, 1994), (2) the bias to attribute events caused externally to specific others rather than situations or chance or *personalizing bias* (PB; Kinderman & Bentall, 1996), (3) the bias to attribute all events to others rather than oneself, regardless of valence (Moritz et al., 2007), (4) the bias to attribute negative events caused by others to intentional hostile motives or the *hostile attribution bias* (HAB; Combs et al., 2007), and (5) aberrant decision making in determining the trustworthiness of ambiguous others, or *trustworthiness bias* (TB; Pinkham et al., 2008).

Individuals from a general schizophrenia sample tend to regard others' intentions as hostile and intentional in ambiguous negative situations (HAB; Combs et al., 2009; Buck et al., 2016b; Lahera et al., 2015; Kanie et al., 2014), and use facial cues differently from controls to make determinations about trustworthiness (TB; Couture et al., 2010; Hall et al., 2004; McIntosh & Park, 2014). Those experiencing active paranoia or persecutory delusions are likely to blame others for negative events and oneself for positive events (EB; Bentall & Kaney, 2005; Combs et al., 2009; Craig et al., 2004; Jolley et al., 2006; Mehl et al., 2010; Randall et al., 2003), attribute the cause of fewer events (regardless of valence) to themselves (SCB; Aakre et al., 2009; Diez-

Alegria et al., 2006; Lincoln et al., 2010; Moritz et al., 2010; Randall et al, 2003; Randjbar et al., 2011), more strongly and consistently attribute ambiguous negative events to others' intentional hostility (HAB; Chang et al., 2009; Combs et al., 2009; Kim et al., 2014) and also systematically regard others as untrustworthy (TB; Buck et al., 2016b, Pinkham et al., 2008; Pinkham, Harvey & Penn, 2016). Further, these biases provide information about clinically important phenomena in schizophrenia, including persecutory delusions or paranoia (Combs et al., 2007b, 2009; Craig et al., 2004; Mehl et al., 2010, 2014; Kinderman & Bentall, 1997b; Langdon et al., 2006; Langdon, Ward & Coltheart, 2010; Lincoln et al., 2010), depressive symptoms and self-esteem (Candido & Romney, 1990; Fraguas et al., 2008; Krstev, Jackson & Maude, 1999; Mancuso et al., 2011; Martin & Penn, 2002; Sanjuan et al., 2009), attachment style (Donohoe et al., 2008), state and trait anxiety (le Gall et al., 2013), and clinical insight (Langdon et al., 2006).

Clinical utility of social cognitive biases

Overall, social cognitive biases appear to provide useful information that differs from the kind of information provided by assessments of social cognitive skills. Social cognitive skills provide information primarily about skill-based functional outcomes like independent living or social functioning. Social cognitive bias measurement is useful in a clinical context as these measures appear to aid in (1) clarifying heterogeneity and identifying subgroups within schizophrenia, (2) identifying treatment targets for cognitive interventions, and (3) providing a window into the cognitive processes that contribute to the emergence of paranoia or persecutory symptoms.

First, each social cognitive bias appears to identify groups within the heterogeneous schizophrenia category. In particular, there is significant variation within a schizophrenia sample with regards to what phenomena create the most significant impediments to functioning, and

how. For example, individuals with primary social cognitive skill impairments may struggle to process incoming messages, remember information, and solve problems, but avoid interpersonal conflicts (Mancuso et al., 2011; Pinkham et al., 2016b). On the other hand, the individual elevated in suspiciousness and hostility may process and remember information effectively, but she or he may avoid interpersonal interaction or approach socializing in a manner that results in conflict (Buck et al., 2016b; Pinkham, Penn & Harvey, 2016). Social cognitive biases provide information underlying domains that might explain or predict the risk of individuals' encountering such outcomes.

Second, social cognitive biases provide an area of for clinical intervention, though large trials examining the effect of interventions have been rare. A few studies (Brakoulias et al, 2008; Tas et al., 2012) have examined the effect of cognitive behavior therapy on externalizing and personalizing biases, but these were not specifically targeting individuals with persecutory delusions. Skills training interventions (e.g. Social Cognition Interaction Training [SCIT; Roberts & Penn, 2009]; Social Cognition Skills Training [SCST; Horan et al., 2009, Horan et al., 2011] and Metacognitive and Social Cognitive Training [MSCT; Rocha & Queirios, 2013]) have taken this approach in challenging participants to adjust their attributions, yet these have had mixed support, with small treatment effects in outpatient and community settings (Hasson-Ohayon et al., 2014; Roberts & Penn, 2009; Roberts et al., 2010; Tas et al., 2012; Kurtz & Richardson, 2012 for a review), moderate treatment effects in inpatient settings (Combs et al., 2007a) and large effects in reducing extreme responding overall (Wang et al., 2013). Overall, meta-analytic work shows small to moderate effects of skills training interventions on attributional style (Kurtz et al., 2016). While considerable work remains in designing and adjusting these interventions, social cognitive biases provide an identifiable target to approach

from a psychosocial perspective. Approaching persecutory delusions from a psychosocial approach can be particularly challenging for clinicians; persecutory delusions (once well-formed) often defy rational counterargument and are maintained despite overwhelming counterevidence (Langdon & Coltheart, 2000). Addressing social cognitive biases might lend themselves to treatment in a way well-formed persecutory delusions do not. Remediating or practicing adjustments to aberrant social cognitive biases might aid clients in preventing or countering potentially delusional beliefs.

Third, from a broad research perspective, the study of social cognitive biases could provide further insight into the processes that underlie psychosis, in particular persecutory delusions. Social cognitive biases are strongly related to these phenomena (Combs et al., 2007b, 2009; Craig et al., 2004; Mehl et al., 2010, 2014; Kinderman & Bentall, 1997b; Langdon et al., 2006; Langdon et al., 2010; Lincoln et al., 2010). A better understanding of these biases might provide groundwork for a cognitive model of persecutory delusions. Findings related to changes in these biases provide testable hypotheses about the cognitions associated with persecutory delusions. Preliminary research suggests that persecutory delusions might emerge as products of dual process interactions of social cognitive biases with stress, time, and context. For example, in their defensiveness model of paranoia, Bentall and colleagues (2001) hypothesized that individuals with paranoia blame others for negative events to reduce associated threats to self-esteem. This hypothesis has been supported by findings demonstrating that paranoid individuals have low implicit and high explicit self-esteem (Mehl et al., 2010; Moritz Werner & Collani, 2006), experience sampling models showing a decrease in mood and self-esteem precedes a spike in severity of persecutory delusions (Thewissen et al., 2011), and findings demonstrating that the (explicit) externalizing bias of individuals with paranoia is more labile than non-patient

controls (Bentall & Kaney, 2005). Further, Moritz and colleagues (2011) found that under a stress induction, individuals with paranoia showed increases in their tendency to attribute responsibility to other people (across event valence), whereas control participants' responses remained the same. While these findings are too preliminary for developed models, they do support a view suggesting that paranoia may result from difficulty regulating the downstream effects of automatic biases (as in Moritz et al., 2011) or an explicit effort to counteract implicit low-self esteem (Lyon et al., 1994). These rich accounts of the emergence of persecutory delusions suggest value exists in treatment approaches that attend to managing stress and challenging biased cognitions when the individual is under threat. Studies of social cognitive biases might allow for that rich process-oriented account of the emergence of persecutory delusions.

Limitations in the study of social cognitive biases

Recent efforts to better understand cognition and functioning in schizophrenia have turned toward large-scale psychometric validation studies. These studies (e.g. MATRICS, Green et al., 2004; SCAF, Green, Lee, & Ochsner, 2013; VALERO, Leifker et al., 2009; SCOPE; Pinkham et al., 2015, 2016b) aim to gather expert consensus on important domains of cognition or functioning, examine available assessment tools from the research literature, create "gold-standard" batteries to examine the area, and validate those batteries for use. In this process, these studies generally recruit large samples of individuals with schizophrenia that are (1) currently medicated, (2) in a stable period of illness, and (3) heterogeneous with regard to presentation (e.g. primary deficit syndrome, primary paranoia, etc.). Using samples of this kind, cognitive assessments are examined (among other attributes) for their ability to (1) distinguish between participants with schizophrenia and controls, (2) reliably provide consistent performance for each

participant (e.g. inter-rater and test-retest reliability), and (3) consistently predict measures of real-world functioning (e.g. independent living skills, role functioning, social functioning) (Pinkham et al., 2015, 2016b).

Such efforts have led to progress in developing theoretical models (Lysaker et al., 2013; Mancuso et al., 2011) and empirically supported instruments (Pinkham et al., 2016b) for social cognitive skills; however, similar success with regard to social cognitive biases has been elusive. A critical development in social cognition research in schizophrenia was the Social Cognition Psychometric Evaluation (SCOPE; Pinkham et al., 2015) study, a large-scale multiphase study gathering expert opinions on core domains and measures of social cognition in this population, establishing consensus through RAND panel methodology (Fitch et al., 2001), and developing, piloting, and validating a gold-standard measurement battery assessing these domains (Pinkham et al., 2016b). In the first phase of the SCOPE study (Pinkham et al., 2015) one bias-related domain was identified as a core domain of social cognition, attributional style/bias, and only one measurement was identified as having adequate psychometric support, the Ambiguous Intention Hostility Questionnaire (AIHQ; Combs et al, 2007c), a measure of hostile attribution bias. Other social cognitive biases did not have sufficient research support, as some biases have not been systematically studied in a clinical context (trustworthiness bias and self-causation bias), and others are inconsistently present in schizophrenia samples, (externalizing bias and personalizing bias; Savla et al., 2013).

However, at the conclusion of the initial psychometric study, because of modest test-retest reliability coefficients and poor relationships of the AIHQ to functional outcomes like independent living skills and role functioning, this measure was removed from inclusion in the final battery. This has raised a contradiction in the literature; attribution bias is regarded as a

well-established and important domain of social cognition with ostensible psychometric support and clinical utility, however, in-depth psychometric review revealed that this measure had limited utility in predicting criterion outcomes, and therefore was excluded from further study (Pinkham et al., 2016b). This complexity has directed current research toward an improved and more nuanced understanding of attributional style, and in particular, the domain studied in SCOPE, the hostile attribution bias.

Hostile Attribution Bias

Attributions, or explanations of the cause of events, can have affective, behavioral, or cognitive consequences for individuals generating them. People differ with regard to the frequency with which they make *certain kinds* of attributions, particularly for important life events. "Attributional style" (AS; or "attributional bias"[1], AB) refers to the *pattern* with which individuals generate explanations of the causes of positive, negative or ambiguous events that occur in their lives (Peterson et al., 1982). The effects of attributional style on emotion and behavior are well-documented in a range of psychiatric disorders, most extensively in mood disorders (Abramson, Seligman & Teasdale, 1978; Sweeney, Anderson & Bailey, 1986), but also in externalizing symptoms in childhood (Crick & Dodge, 1994), anxiety disorders (Heimberg et al., 1989), eating disorders (Mansfield & Wade, 2000), and pediatric chronic illnesses (Schoenherr et al., 1992).

Initial work on attributional style in patients with schizophrenia merely examined these individuals as a psychiatric control sample compared to those with depression, who demonstrated a "depressive attributional style" (Abramson, Seligman & Teasdale, 1978). Depressed individuals are more likely to make internal attributions about negative events (i.e.,

[1]These terms will be used interchangeably in the current review.

[2] This was after a Bonferroni connection to account for the number of correlations computed in the study (p = .05 / 45).

rather than to others or circumstances), and to regard these characteristics as stable (i.e., unchanging) and global (i.e., relevant across domains of life). People with schizophrenia, on the other hand were significantly *less* likely to attribute negative events to themselves, not only in comparison to the depressed subjects, but also to non-clinical controls (externalizing bias). While these initial observations raised important questions about the nature of attributions in psychosis, later developments demonstrated inconsistent results in samples with schizophrenia and persecutory delusions (Savla et al., 2013).

However, future research noted that even when events are clearly negative and attributable to others, individuals still differ in patterns with which they interpret others' intentions. Combs and colleagues (2007c) argued that individuals with schizophrenia might be more likely to interpret others' actions as intentional and hostile, and this tendency may contribute to paranoia. This – the hostile attribution bias – describes the tendency to believe negative events are the result of intentional, hostile actions, rather than accidents or chance. Combs et al. (2007c) presented a model of three related biases in these negative situations: *blame* (the tendency to think others acted intentionally), *hostility* (the tendency to ascribe hostile motives) and *aggression* (the tendency to describe a hypothetical aggressive behavioral response to the event). They hypothesized that because these ambiguous situations are especially difficult to interpret for those with social cognition impairments, this ambiguity and challenge would result in a tendency to interpret others' actions in a hostile manner. To examine these biases, they presented participants with hypothetical ambiguous scenarios involving the actions of others, and asked them to identify the extent to which the actions of others were purposeful, blameworthy, and made them angry. This procedure was developed as a brief instrument, the Ambiguous Intentions Hostility Questionnaire (AIHQ; Combs et al., 2007c).

The Ambiguous Intentions Hostility Questionnaire (AIHQ)

The rationale of the AIHQ is that as individuals with schizophrenia show social cognition impairments, they are less apt to correctly attribute the causes for others' behavior, and in the absence of information, may show a bias toward threat (Combs et al., 2007c). Each item of the AIHQ presents individuals with ambiguous scenarios (e.g. "You walk past a bunch of teenagers at a mall, and they start to laugh") and asks three questions with ratings provided on a Likert scale: "did [the person(s)] do this to you on purpose?", "how angry would it make you feel?", and "how much would you blame [the person(s)]?". The participant also generates an attribution for why the event occurred, and how they would hypothetically respond behaviorally. The Likert scale questions are summed to calculate a *blame score* and the open-ended responses are evaluated by an independent rater (also on a Likert scale) for how hostile the attribution and how aggressive the hypothetical behavioral response. Thus, the *hostility bias* describes the tendency for an individual to interpret an other's action as ill-wishing or hostile, the *aggression bias* describes the tendency for the individual to respond to ambiguous situations with antisocial or aggressive behaviors (e.g. to shout at others), and the *blame score* is an index measure combining judgments of blame and feelings of anger. Each scenario item is varied by its ambiguity, and these items are separated into three categories: accidental (other actors seemingly acting unintentionally), intentional (other actors seemingly acting intentionally), and ambiguous (intentions of actors are unclear).

In a large sample of non-clinical controls, the AIHQ demonstrated strong inter-rater reliability (for hostility and aggression scores) and internal consistency. To validate the measure, Combs and colleagues (2007c) used hierarchical linear regression to examine whether the AIHQ predicted paranoia beyond other extant predictors of paranoia. They found that indeed the AIHQ

(combined Aggression, Hostility and Blame scores in ambiguous situations) predicted paranoia above and beyond the influence of gender, ethnicity, extant attributions measures, Paranoia-Suspiciousness Questionnaire Hostility subscale, and psychosis proneness as measured by the Chapman Perceptual Aberration Scale. Combs et al. (2007c) demonstrated an advantage for the AIHQ over previous measures of personalizing blame, which were uncorrelated with these outcomes in the same sample.[2] The blame score across all intentionality scenarios correlated significantly with paranoia and hostility scales; only Hostility Bias in ambiguous scenarios and Aggression Bias in accidental scenarios correlated with any other convergent predictors (self-reported paranoia). Thus, the authors concluded that validity analyses were strong for the Blame scale, mixed for the Hostility bias, and weak for the Aggression bias. Therefore in subsequent research, greater attention has been paid specifically to the blame and hostility subscales.

Empirical support for hostile attribution bias

Group differences

Group differences exist between schizophrenia and control groups in hostile attribution bias, as individuals with schizophrenia appear to have elevations in blame and hostility (Buck et al., 2016b, 2017; Kanie et al., 2014; Lahera et al., 2015). These elevations appear most extreme among patients with persecutory delusions. Combs and colleagues (2009) demonstrated that patients with persecutory delusions (and schizophrenia) had elevations on all AIHQ subscales compared directly to both participants with schizophrenia without persecutory delusions and non-clinical controls. Interestingly, non-persecutory deluded participants with schizophrenia looked statistically indistinguishable from non-patient controls. When examined with correlational analyses, the AIHQ has demonstrated consistent relationships to paranoia in non-

[2] This was after a Bonferroni connection to account for the number of correlations computed in the study ($p = .05 / 45$).

patient control samples (Combs et al., 2007c; Combs et al., 2013), paranoid individuals with a range of diagnoses (Combs et al., 2009), individuals at ultra-high risk (An et al., 2010), as well as combined across phases of psychosis (Kim et al., 2014). It appears that these consistent relationships are not specific to self-report questionnaires or clinical interviews as well. Relationships of hostile attribution bias to paranoia, hostility and suspiciousness persist when assessing them in a number of ways, including objective personality test subscales (Combs et al., 2009), and self-report questionnaires (Chang et al., 2009; Combs et al., 2007b; Combs et al., 2009; Kim et al., 2014), interview-based ratings of hostility/suspiciousness symptoms (An et al., 2010; Mancuso et al., 2011), as well as in-the-moment ratings of anger (Kim et al., 2014). Only one study (Buck et al., 2016b) demonstrates any prospective relationships, demonstrating that hostile attribution bias predicts prospective worsening in emotional discomfort and hostility symptoms. Further, different from other social cognitive biases (EB, Delvyder et al., 2013; Humphreys and Barrowclough, 2006; Langdon et al., 2013); PB, Delvyder et al., 2013), an increased hostile attribution bias presents in early psychosis as well. Hostility (An et al., 2010) and blame scores (le Gall et al., 2013) are elevated in first-episode psychosis compared to non-clinical controls.

Relationship to non-psychosis psychopathology

In schizophrenia, early psychosis and risk-states, comorbid anxiety and depressive symptoms are common (Majadas et al., 2012; Wassink et al., 1999). Complicating models of attributional style in psychosis is the fact that attributional biases are predicted by levels of other symptoms of psychopathology as well; depression ratings predict AIHQ scores modestly in a combined patient/non-clinical sample (Combs et al., 2009), and highly in combined high-risk and early psychosis sample (Kim et al., 2014), a sample in which anxiety also significantly

predicts AIHQ as well (Kim et al., 2014). Le Gall and colleagues (2013) showed that for first-episode psychosis patients, blame and hostility biases were associated with high anxiety and low self-esteem. A combined AIHQ total score was strongly correlated with not just positive symptoms of psychosis, but anxiety and mood symptoms as well (Mancuso et al., 2011). This raises questions about the specificity to psychosis of this bias, or if instead it is related generally to psychopathology more broadly.

No studies have compared a depressed or anxious sample to a schizophrenia sample in AIHQ scores, but other statistical techniques have been employed to clarify these relationships. An and colleagues (2010) examined the effect of controlling for depression on the relationship between the AIHQ and the paranoia scale, finding that for individuals in the midst of a psychotic episode, controlling for depression *increased* the relationship between AIHQ scores (both blame and hostility) and the paranoia scale, but for individuals at ultra high risk for development of psychosis, controlling for depression decreased it, almost fully accounting for it. They also noted that individuals at ultra high risk had significantly lower self-esteem than non-clinical controls and patients in the midst of a first psychotic episode. This pattern of results seem to suggest that – as is the case for externalizing bias – subclinical at-risk or sub-threshold individuals might present with hostile attributional biases that are more closely related to depressive symptoms than positive symptoms of psychosis. This would be indicative of a qualitative difference between those with psychosis and those at risk (rather than a continuous dimensional change), and that these attributional biases are sustained independently of mood symptoms in later phases of the disorder.

It might also be possible that measurement artifacts could be contributing to conflation of depression and hostile attributional style. Several items of the AIHQ seem to overlap

conceptually with depressive cognitions. For example, the individual who responds with a hostile attribution to the item "You walk past a group of teenagers and they start to laugh" (e.g. "they were laughing at how I look") might suffer primarily from paranoia, or they might just be affected by cognitions associated with depression, given the overlap of psychosis and dysphoria (Lako et al., 2012; Majadas et al., 2012). Thus, studies comparing individuals with psychosis to controls or examining relationships of attributional style to functioning could be impeded by not employing models that account for the role of depression and anxiety in affecting both variables. Granted, these relationships may not be mutually exclusive (i.e., it's possible for hostile attributions to be relevant in both depressed and paranoid people) but the coexistence and confounding of these patterns leaves questions unanswered about the specific clinical relevance of these patterns in cognition to psychosis.

Relationship to functioning

To match findings demonstrating consistent relationships of skill-based social cognition scales (e.g. emotion perception, theory of mind) to functioning (Fett et al., 2011 for a meta-analysis), research from the last decade has examined whether performance on the AIHQ is similarly related to functioning and living skills. Few have demonstrated relationships; but there is some evidence of relationships of the AIHQ with clinician-rated GAF scores (Lahera et al., 2015), executive functioning (Kim et al., 2014), outdoor activity (Elnakeeb et al., 2010) and social engagement (Kanie et al., 2014). More studies, however, have demonstrated a lack of relationships to these outcomes. Mancuso and colleagues (2011) showed that a combined factor-derived AIHQ total score (combining hostility, blame, and aggression) showed no relationships to measures of functioning, including a skills-based living skills measure, and interview-rated measures of independent living, social engagement, and work skills. In the ongoing large-scale

SCOPE study (Pinkham et al., 2016b), the same pattern emerged (no correlations to the subscales of the Specific Levels of Functioning [SLOF], and the UCSD Performance-Based Skills Assessment [UPSA]). As mentioned, it was this pattern of findings (or lack thereof) that led to the exclusion of the AIHQ from the SCOPE gold-standard battery of social cognition in schizophrenia.

A hostile attribution bias may, however, have a more significant relationship to functioning in areas related to interpersonal conflict. Waldheter et al. (2005) showed that hostile attributional bias predicted violence on an inpatient unit both concurrently and prospectively. Similarly, a re-evaluation of SCOPE data according to interpersonal conflict criteria (Buck et al., 2016b) reveals that the AIHQ is more closely related to interpersonal conflict (e.g. physical and verbal fights with others) and personality ratings of hostility and paranoia than independent living or work skills. Findings in this area are scant, but patterns thus far suggest that hostile attributional style is a factor that might affect one's propensity to engage in conflicts, both physical and verbal, and that this may be a more specific pathway through which attributions could affect functioning.

Limitations and persisting questions

Process orientation

Existing measures of hostile attribution bias provide access to the outcomes of biased social judgments. They do not provide information about the processes one engages in order to arrive at such judgments. Paradigms from social psychology have demonstrated the process through which all individuals generate social judgments often involves two sets of related processes, automatic and controlled influences (Payne, 2001). While automatic processes are efficient, involuntary, immediate, and operate outside the individual's awareness, controlled

processes describe effortful, controllable and conscious. According to such a model, cognitive processes are comprised of automatic reactions that can be accepted, rejected, or amended "downstream" moments later with cognitive processing. Thus, cognitive processes (and particularly biases) are not the product only of immediate reactions, but one's ability to regulate and control such initial reactions.

While this has not been examined in depth, there is some reason to believe that models of hostile attributions in psychosis may be better described with a similar dual process model of automatic and controlled processes. Preliminary research suggests that individuals with paranoia may have difficulty regulating automatic biases under stress (as in Moritz et al., 2011, 2015), may be more susceptible to primes (Hooker et al., 2011) or may present opposite biases across implicit and explicit conditions (Lyon et al., 1994). Existing hostile attribution bias measures ignore such processes. For example, the large-scale SCOPE study (Pinkham et al., 2015, 2016b) examined only comparisons of mean scores across a clinical and control sample as well as correlations with symptom and functioning outcomes. This does not address questions about how individuals make social judgments when in acute states of illness, experiencing high levels of stress, or in a complex social situation more ecological valid to the social environment. And indeed, such process-oriented models of social cognition are regarded as a "new frontier" for research in the area. As described by Roberts and Pinkham (2013) in their review of future directions for the study of social cognition in schizophrenia, "the dual process framework provides a strong basis for applying social psychological principles to the study of social cognition in schizophrenia," particularly in order to distinguish between "(1) diminished controlled processing capacity, and (2) excessively salient and aberrant automatic social cognitive impressions (p. 409)."

Theoretical model

Another limitation of the hostile attribution bias is how it ostensibly combines several constructs into one assessment. Specifically, in asking individuals to ascribe levels of blame and hostility for a range of negative events, this measure confounds immediate biases about negativity of others' actions with beliefs about the intentionality of others' actions. The scenarios of the AIHQ are ambiguous and involve negative outcomes (e.g. a friend doesn't attend a dinner, an important person skips an appointment). In this way, it cannot be determined whether individuals have an increased bias toward attributing intentionality in negative events specifically, or in all events. Using other instruments, individuals with schizophrenia appear to assign intentionality to neutral actions (e.g. "he set the alarm off" being perceived as intentional rather than accidental) at higher rates than non-patient controls (Peters et al., 2014; Peyroux et al., 2014), though these results have limited success in predicting symptoms or functional outcomes.

It is also unclear if hostile attributional style is specific to the emergence of paranoia, or if it emerges as a result of (or concurrently with) other comorbid symptoms like anxiety and depression. While it appears hostile attributional bias is related to these other forms of psychopathology (Combs et al., 2009; le Gall et al., 2013; Kim et al., 2014; Mancuso et al., 2011), it is possible that in the absence of depression, this bias might be a particularly strong indicator of psychosis (An et al., 2010) or that there may be a more complex relationship according to which this thinking affects the presentation of other symptoms (e.g. ideas of reference, Morrison & Cohen, 2014). Finally, while particularly robust findings suggest that hostile attributional style has little impact on independent living skills, continued work should

explore the range of more appropriate criterion validity outcomes that might be impacted by this bias, e.g., interpersonal violence (Waldheter et al., 2005) or other conflict (Buck et al., 2017).

Measurement issues

There are several flaws that were first observed in the initial study of the AIHQ (Combs et al., 2007c). First, the subscales of the AIHQ are of unclear incremental validity. As mentioned, the aggression bias was regarded as having weak psychometric support initially (Combs et al., 2007c) and subsequent research has supported this, as it has not demonstrated relationships to paranoia scales (Combs et al., 2007c), hostility (Combs et al., 2009) and trait anger questionnaires (Jeon et al., 2013), and the hostility and blame subscales of the AIHQ (Buck et al., 2017; Combs et al., 2009, Mancuso et al., 2011). Factor analytic work (Buck et al., 2016a) has demonstrated a small factor loading with the other subscales of the AIHQ as well. One reason for the unimpressive results of the aggression subscale might be its lack of variability, as reporting an aggressive action (e.g. a plan to fight an individual that slights the participant) is a relatively infrequent response or subject to social desirability bias (Buck et al., 2017). On the other hand, it is also unclear the extent to which the hostility and blame subscales diverge from one another, or if indeed they are two items measuring the same underlying construct. Previous research suggests that the two scales are highly correlated with one another, both in psychotic disorder and control samples (Buck et al., 2017; Combs et al., 2009, Mancuso et al., 2011).

Second, the AIHQ has few items. It is possible that this limits its sensitivity and power to predict important outcomes or distinguish between clinical and control groups. Collecting a high number of observations in assessments of social cognitive biases is particularly important, as these biases reflect a general tendency to make social determination in certain ways across situations and contexts. The hostility and aggression biases involve open-ended responses that

are later rated by trained researchers (Combs et al., 2007c). This characteristic – because of its burden on providers and time – potentially limits the number of items, reducing variance in measure performance and therefore statistical power. This is concerning given the fact that the blame scale of the AIHQ sufficiently relates to interpersonal conflict outcomes without the help of the additional subscales (Buck et al., 2016b, 2017).

Third, biases in social judgments emerge across a range of situations that vary according to how obvious or demanding their cues. For example, a situation where intention appears less ambiguous (e.g. "a person cuts you off in traffic") may result in a hostile attribution for a higher number of participants from both clinical and normative samples. A more ambiguous action, on the other hand (e.g. "you walk by a group of teenagers and they start to laugh") might generate varied responses, the patterns in which might relate to clinical factors. Because the initial large-scale analogue sample validation study showed highest convergent correlations for the ambiguous items (Combs et al., 2007c), most subsequent studies only include these items to simplify administration and reduce testing time. While this did successfully simplify the scale, it potentially limits variability by both lowering the ceiling of the scale and raising its floor. And in fact, work in progress demonstrates that the AIHQ accidental items provide incremental validity above and beyond totals of the ambiguous items in predicting independent living skills and informant-reported levels of functioning (Buck et al., 2017). Thus, it should be noted that a planned and systematic varying of the demand of the situational cues might provide a more sensitive and holistic measurement of hostile attribution bias.

Finally, and perhaps most importantly, as mentioned, social cognitive biases describe phenomena that differ from skills in important ways that affect their measurement. While adaptive functioning in social cognitive skill consists of being skilled (e.g. the *more skilled* the

better), social cognitive biases differ. In interpreting others' actions as hostile or intentional, it would neither be adaptive to *always* assume hostile intentions or to never do so. Rather, adaptive functioning in a social cognitive bias ought to be defined in relation to an adaptive level of hostile attribution according to the situational context of each item. Leading measures of hostile attribution bias in schizophrenia (AIHQ; Combs et al., 2007c) do not do this. Comparing mean values between groups allows for comparison of raw values, but it does not characterize whether the individual is demonstrating elevated responses because of an inability to understand situational cues (i.e. low discernibility) or an actual bias toward hostility (i.e. true bias). More sophisticated models can compare attributional responses against *expected* responses such that this fine-grained analysis can be conducted.

Directions for future research

The limitations identified in the previous literature on social cognitive biases in schizophrenia provide directions for continued research in this area. First, there is a need for paradigms that examine the processes through which biased judgments emerge. In particular, research should examine the extent to which social cognitive biases emerge through a dual process with both automatic (implicit) and controlled (explicit) processes. This could provide greater insight into determining whether individuals with hostile attribution bias appear to have an automatic bias, a difficulty regulating or controlling a downstream bias under stress (as suggested by Moritz et al., 2011), or some combination of both. Second, there are a number of questions related to the theory underlying the hostile attribution bias in psychosis. In particular, because the AIHQ subsumes overall judgments of intentionality and judgments of intentional negative action, it does not clearly discriminate those who tend to judge all actions as intentional from those with a specific bias to interpret others' actions in a hostile manner. Preliminary

research suggests people with schizophrenia do tend to interpret others' actions as more intentional than non-patient controls (Peyroux et al., 2011) but this measurement was not conducted in a manner allowing examination of automatic or controlled processes. Third, there is a need to test potential improvements to existing measures of hostile attribution bias, in particular increasing the number of items and varying them with regard to a criterion variable (e.g. ambiguity, valence) that pulls for the phenomenon of interest (e.g. a hostile interpretation of motives, externalizing blame). One specific way to address these limitations is to turn toward the non-patient control literature, and in particular innovative measurement technologies from social psychology. These paradigms provide potential answers to these limitations.

Paradigms from social psychology

Two particular biases from the social psychology literature studied in large non-patient control samples address these limitations, the intentionality bias (Rosset et al, 2008) and the immorality bias (Hester, Payne & Gray, 2017). Each of these assessments has been validated in a non-clinical population and addresses the identified limitations in the hostile attribution bias in schizophrenia by (1) examining automatic and controlled processes of automatic social judgments with (2) a higher number of items that (3) systematically vary with regard to the criterion variable. Specifically, the intentionality bias provides estimates of automatic and controlled processes that underlie judgments of intentionality. The immorality bias task provides estimates of automatic and controlled processes in the individual's tendency to specifically view others' actions as evil or nefarious. This allows for a distinction in these judgments that cannot be made with the AIHQ in its current form. Each of these assessments may tap into the cognitive domains that tend to go awry in psychosis. Additionally, the intentionality and immorality biases are two biases that are *normative* to emerge among non-clinical individuals under time pressure

or cognitive load. Examining these measures could allow for a more precise differentiation between the biased judgments of non-patient controls and participants with schizophrenia. In particular, applying these paradigms to a clinical sample allows for researchers to determine whether the social cognitive biases of individuals with schizophrenia or paranoia are simply an exaggeration or unchecked expression of a bias shared by non-clinical samples. This contribution could lay the groundwork for a normalizing cognitive model of the gradient between normative and pathological thinking associated with psychosis.

Intentionality bias

When interacting with others, one important determination in response to both positive and negative actions is whether the individual in question regards the other's action as intentional. Philosophers and cognitive scientists have understood one's intentionality of actions as being characterized by the individual desiring an outcome, believing an action will lead to the outcome, an intention to perform this act, and awareness of performing the act. Non-clinical samples generally tend to agree about overall determinations of intentionality (Malle & Knobe, 1997).

However, while conceptually, people tend to agree on their notions of intentionality, there exist individual differences with regard to how frequently individuals tend to regard others' actions as intentional. This can have significant downstream effects on moral judgments (Knobe, 2008), emotions, and behavioral responses (Wegner, 2002). For example, the individual that interprets a friend's ignoring her or him at a social gathering as an intentional "snub," rather than an accidental oversight is more likely to experience anger or hostility in response.

Recent research has examined the processes underlying these kinds of judgments, as adults tend to have a bias to interpret ambiguous actions as intentional rather than accidental.

These data suggest that that judgments of intentionality are best understood in terms of a dual-process model demonstrating that these judgments are the product of (1) immediate automatic judgments as well as (2) more complex or cognitively demanding attempts to revise or override these initial quick judgments. Specifically, participants tend to show a bias toward interpreting more events as the result of intentional action when under time pressure. The ability to identify actions as accidental results from the "ability to override, and thus inhibit" the tendency to view "everything anyone ever does as intentional" (p. 772; Rosset, 2008). This ability is hypothesized to emerge through the ability to "override an initial interpretive impulse" and consider additional information including social norms, others' beliefs, goals, and intentions.

Rosset and colleagues (2008) have demonstrated this bias through the use of their intentionality bias task, which presents participants with a number of sentences, varied according to their pre-rated intentionality (based on pre-collected and untimed ratings), featuring unambiguously accidental (e.g. "he stubbed his toe"), unambiguously intentional (e.g. "he threw the football), prototypically accidental (e.g. "the girl popped the balloon"), and prototypically intentional (e.g. "he took an illegal left turn") items. Participants are asked to determine whether these items describe actions that were done "on purpose" or "by accident," and do so in several (counterbalanced) task blocks, varying whether individuals are required to respond under time pressure (2400 ms), and another with reduced time pressure (5000 ms). Participants were significantly more likely to regard unambiguously accidental and prototypically accidental items as intentional in the speeded condition, while estimates for the unambiguously intentional and prototypically intentional remained the same.

Only one study has examined this form of intentionality bias in schizophrenia. Peyroux and colleagues (2014) presented a group of participants with schizophrenia and non-patient

controls with 72 sentences that varied with regard to whether they were previously rated as clearly intentional or clearly unintentional, and indeed, individuals with schizophrenia were more likely than the control participants to rate all the sentences as the result of intentional action rather than by accident. Performance also was significantly related to scores on clinician-rated symptom interviews, including poor impulse control and excitement. Importantly, however, this study did not examine the ability of all participants to regulate initial responses, but rather gave all items without time pressure.

The intentionality bias is a viable candidate for improved measurement of social cognitive biases in schizophrenia for a number of reasons previously mentioned. First, it examines a cognitive bias that appears to present in continuous levels between non-patient and clinical samples. Second, the intentionality bias task provides a range of items of varying situational demands; some items are designed to be interpreted as clearly intentional, others clearly accidental, and the remaining items of varying clarity with this criterion. This, along with a higher number of items, allows for a more fine-grained examination of subtle biases in intentionality judgments. Finally, the intentionality bias as designed by Rosset (2008) allows for an examination of the process of intentionality judgments, by comparing bias levels across differing time pressure conditions.

Immorality bias

A second bias present in a normative population also is a strong candidate for the understanding of social judgments in schizophrenia. Hester and colleagues have proposed that when tasked with making a moral judgment about the ambiguous actions of others, individuals demonstrate an automatic "*immorality bias*" (p. 8, Hester, Payne & Gray, 2017), meaning that immediate or automatic judgments are more likely to view others' intentions as evil or nefarious.

This bias appears to be related to or draw from components of three other biases consistently demonstrated in non-clinical samples, in particular a bias to detect threat (Cisler & Koster, 2010), to overvalue negative information (Hilbig, 2009; Skowronski & Carlston, 1989), as well as the aforementioned intentionality bias (Rosset, 2008). However, the bias to assess ambiguous actions as immoral appears to result from an emergent process drawing upon each of these components yet is greater than the sum of all three. Morality judgments demonstrate a statement of global social beliefs about the likelihood of others' intentional, nefarious, or harmful behavior (Gray, Schein & Ward, 2014). Further, these judgments are inherently specific to social interactions with others, as morality tends to apply to communities of individuals interacting with one another.

Hester and colleagues presented a sample of 28 sentences, stratified into categories – prototypically immoral (average approximately 75% immoral) and prototypically nonmoral (average approximately 25% immoral) – based on immorality ratings previously collected in a pilot study of 60 non-clinical participants providing responses without time pressure. In a similar paradigm to Rosset's intentionality bias task, participants were asked to identify each sentence as "immoral" or "not immoral," in two time pressure conditions (1500 ms and 5000 ms). The pre-rated categorization of the items allowed for the use of process dissociation (Jacoby, 1991). Process dissociation is a mathematical procedure in which participant responses are compared to pre-set criterion ratings to determine the patterns in consistency with which participant responses adheres to the pre-set criterion. This allows estimates that differentiate automatic and controlled processes. Automatic processes are not consciously initiated and independent of cognitive control, whereas controlled processes recruit effortful conscious cognition. With reduced ability to use cognitive control (e.g. under time pressure or cognitive load), individuals are more likely

to rely on automatic processes. Hester and colleagues (2017) demonstrated that indeed, under time pressure, participants demonstrated a reduced ability to engage controlled processes, and thus were more likely to regard items as immoral, and particularly with regard to the prototypically non-moral items. This was interpreted ultimately that "the immorality bias represents a stable preference for assumptions of wrongdoing, and that controlled processing can be used to override this initial preference when cognitive resources are available." (p. 29; Hester, Payne & Gray, 2017).

The present study

The proposed study applies these innovative research paradigms – the intentionality bias and the immorality bias – to the study of hostile attribution bias in schizophrenia. The implicit measurement paradigm and process dissociation procedure provides several quantifiable outcomes, one for each time condition (low or high time pressure): *total score*, or the total number of items identified as immoral or intentional, as well as the extent to which participant responses correlate with criterion standards (inherent likelihood ratings, or ILRs) for responses (*discernibility score*, regression slope) and the extent to which participant responses are biased in either direction (*pure bias score*, regression intercept). Additionally, using a process dissociation procedure, this paradigm produces two additional values, an *automatic bias estimate*, or an estimate of the automatic bias to see acts as immoral or intentional, and *control estimate*, or an estimate of the extent to which one's pattern of responding is based upon effortful cognitive processing. These values provide the estimates necessary to address a number of aims and hypotheses related to the automatic and controlled components of hostile attribution bias judgments in schizophrenia.

Aims and hypotheses

Aim #1. Construct validity, group differences

The first aim of the present study is to compare the levels of each of these biases in schizophrenia by comparing *total scores*. With regard to the immorality bias, while a general population demonstrates a bias to view ambiguous events as immoral under pressure (Hester, Payne & Gray, 2017), several factors suggest a schizophrenia sample will demonstrate further elevations compared to controls. Previous work on hostile attribution bias demonstrates those with schizophrenia are more likely to view ambiguous actions as hostile and intentional (Combs et al., 2007c; 2009), and the immorality bias appears face valid with this construct. With regard to the intentionality bias, previous research has demonstrated increased intentionality bias amongst a heterogeneous sample of individuals with schizophrenia (Peyroux et al., 2013). This has not been conducted, however, with a dual-process paradigm. Hypothesis: It is hypothesized that individuals with schizophrenia will make more attributions of intentionality (intentionality bias) and evil (immorality bias) than non-clinical controls.

Aim #2. Automatic and controlled processes

Individuals with schizophrenia present with elevations in the tendency to make hostile interpretations of others' actions (Combs et al., 2007; 2009) as well as deficits in processing speed (Kern et al., 2011) and executive functioning (Green et al., 2000). As intentionality bias (Rosset, 2008) and the immorality bias (Hester, Payne & Gray, 2017) are present in pressured judgments of non-patient controls, it is possible that reports of biased social judgments amongst individuals with psychosis result from difficulty regulating or correcting initial biased judgments similar to a normative sample. This might point to a process of an initial biased judgment as well as failure to correct or adjust as a result of neurocognitive impairment. Hypothesis: It is hypothesized that participants with schizophrenia will demonstrate higher automatic bias across

time conditions and reduced controlled processing relative to controls across condition, and this difference will be amplified during the slow condition.

Aim #3. Psychometric analysis.

The third aim is to examine the utility of each task as an indicator of relevant clinical phenomena. Previous research suggests that hostile attribution bias (Combs et al., 2009) is related primarily to paranoia, hostility, persecutory delusions, and interpersonal conflict outcomes (Buck et al., 2016b, 2017; Combs et al., 2007; 2009) as well as predict change over time in hostility symptoms (Buck et al., 2016b). Hypothesis: It is hypothesized that intentionality and immorality biases will demonstrate convergent (concurrent relationships to hostility, positive, and suspiciousness/persecution symptoms), criterion (relationships to trait paranoia, trait hostility, observed hostility, aggressive behavior, and interpersonal conflict outcomes) and discriminant validity (lack of relationships between bias estimates and neurocognition).

Exploratory Aim #1. Incremental validity

Previous research suggests that measures of social cognitive bias provide additional information about hostility, paranoia and interpersonal conflict that is not captured by social cognitive skill measures (Buck et al., 2016a). Given that these measures are proposed to be improvements on assessments of these social cognitive biases, we will also examine the extent measures of intentionality and immorality bias predict hostility, paranoia, and interpersonal conflict above and beyond the SCOPE social cognitive skills battery.

Exploratory Aim #2. Relationships beyond Neurocognition

In our first exploratory aim, we will examine whether differences in control estimates are related to deficits in neurocognition. This follows from assumptions of dual process paradigms that the ability to control judgments stems from effortful cognitive processing (Payne, 2001),

including executive functioning and processing speed. This exploratory aim suggests that while bias estimates should be unrelated to neurocognitive functioning, the ability to control the expression of these biases may be specifically related to it.

Exploratory Aim #3. Relationships to social functioning, social skills

Previous measures of social cognitive bias have inconsistent or low relationships to overall measures of social skill and social functioning (Buck et al., 2016a; Mancuso et al., 2011). There are multiple explanations for this pattern of results, among them either the possibility that (1) hostile attribution bias measures are limited, or (2) hostile attribution bias is relevant only to paranoia, hostility, and interpersonal conflict outcomes. With the proposed improved measurements available in the present study, we seek to examine these explanations. Thus, we will also examine the relationships of each pure bias estimate with role functioning, social functioning and social skill.

Exploratory Aim #4. Relationships to non-psychosis psychopathology

Previous research suggests that hostile attribution bias is related not only to paranoia and persecutory delusions, but also mood and anxiety symptoms. In order to explore the specificity of these biases to paranoia, we will also examine the relationship of these biases to non-psychosis psychopathology, particularly cognitive, emotional discomfort, and negative symptoms

CHAPTER 2: METHODS

Sample

Data collection was completed in the final phase of the SCOPE study (Pinkham et al., 2015, 2016). Number (n = 217) participants with schizophrenia or schizoaffective disorder and number (n = 154) controls were recruited to three different research sites, the University of North Carolina at Chapel Hill, The University of Texas at Dallas and the University of Miami Miller School of Medicine. Participants had diagnoses (schizophrenia or schizoaffective disorder) confirmed with a structured clinical interview administered by trained research assistants. Participants with schizophrenia were included in the study if they were not hospitalized in the previous two months, were stable on one medication for at least six weeks, and had no changed in dose in two weeks. Participants in both groups were excluded if they met any of the following exclusion criteria: 1) current or past pervasive developmental disorder, 2) low IQ (< 70), 3) current or past medical or neurological conditions that may affect participation, 4) presence of sensory limitations, 5) presence of substance abuse in the past month, or 6) presence of substance dependence not in remission for at least 6 months. The full sample completed the IBT as well as the remainder of the proposed study battery. However, due to study protocol limitations and time constraints, the immorality bias task was only administered at one site (University of North Carolina at Chapel Hill), meaning that only a subset of participants (n = 42 in the schizophrenia group; n = 58 in the control group) completed this task.

Measures

Social cognitive biases

Intentionality Bias Task (Rosset, 2008)

A slightly modified version of Rosset's (2008) intentionality bias procedure was used to assess bias toward intentionality. This assessment presents participants with sentences varied according to their pre-rated intentionality (based on previously collected and untimed ratings), featuring possibly intentional (average of 25% of participants rating statement as intentional; e.g. "the girl popped the balloon"), and probably intentional (average of 75% of participants rating statement as intentional; e.g. "he took an illegal left turn") items. Participants are asked to determine whether these items describe actions that were done "on purpose" or "by accident," and do so in several (counterbalanced) task blocks, varying whether individuals are required to respond under time pressure (2400 ms), and another with reduced time pressure (5000 ms). Comprehension of the task and definition of anchors (i.e. "on purpose, "by accident") was confirmed by trained research assistants by asking participants their understanding of the task. Participants completed 14 practice trials, 14 trials with low time pressure, and 14 trials with high time pressure. This methodological choice was made given previous findings suggesting that when participants complete high time pressure conditions first, they continue strategies employed in the first set of items and answer all items quickly, thus invalidating the manipulation between conditions. Raw scores were totaled as the number of "on purpose" responses across time pressure conditions (ranging 0 to 28).

Immorality Bias Task (Hester, Payne & Gray, 2017)

The final version of Hester's (Experiment 7; Hester, Payne & Gray, 2017) short sentence categorization task was used to examine the immorality bias. All items were previously rated (without time pressure) by a general population sample according to whether they are "not

immoral" or "immoral," resulting in two categories of items, characterized as "prototypically immoral" (average of 75% of participants rating statement as immoral) or "prototypically not immoral" (average of 25% of participants rating statement as immoral). These two categories (14 items in each category) are presented to participants in two blocks, one block with high time pressure (1500 ms to respond) and another with low time pressure (5000 ms to respond). Participants will be asked to respond whether they found the act to be "immoral" or "not immoral." Again, trained research assistants confirmed comprehension of the task and definition of anchors (i.e. "immoral," "not immoral"), and for the same reason as for the intentionality bias, the low time pressure trial block preceded the high time pressure block. Raw scores were totaled as the number of "immoral" responses across time pressure conditions (ranging 0 to 28).

Discernibility (slope) and pure bias (intercept) estimates

In previous pilot testing for both the intentionality (Rosset et al., 2008) and immorality (Hester, Payne & Gray, 2017) assessments, data were collected on a non-clinical sample responding to each item without time pressure. These values represent how an individual would be *expected* to respond with normative functioning and under normal circumstances. These values were called "intentionality likelihood ratings" (p. 774, Rosset, 2008) when referring to the intentionality bias. For consistency across intentionality and immorality biases; in the present study we will call these values *inherent likelihood ratings* (ILRs). These values allow a determination of whether individuals respond differently than criterion values because they are either (1) not attending to items or discerning the target variable, or whether they are (2) demonstrating a pure bias that reflects a greater likelihood to interpret actions as intentional or immoral. Thus, in addition to raw total values, the present study will examine the extent to which each participant response adheres to the pre-set criterion. In order to do this, each individual's

responses (binary yes or no responses) are regressed on the percentage values collected from pilot testing on each task (Hester, Payne & Gray, 2017; Rosset et al., 2008). These regressions generate two values, a slope (β) or *discernibility score* – indicating the extent to which individuals' responses differ across items that adheres to the non-pressured non-clinical criterion response pattern – as well as an intercept, or *pure bias estimate* – indicating the extent to which the respondent tends to provide responses that are elevated in perceived immorality or intentionality. These values were collected within each participant, within each time pressure condition, such that each individual has a discernibility and bias estimate for each the slow and fast conditions for each task.

Process dissociation procedure

Process dissociation (Jacoby et al., 1993) is an algebraic manipulation that allows estimates for the automatic and controlled processes that underlie quick judgments. In such judgments, responses indicate presence of a target phenomenon (i.e. an immoral act) or its absence (i.e. a non-immoral one). Control is defined as the ability of the individual to "produce a particular response when they intend to, but not produce the response when they intend not to" (p. 183, Payne, 2001). When applied to the present study, a similar logic can apply, assuming the previously demonstrated automatic biases of non-clinical samples toward attributing evil and intentionality, using Payne's (2001) previous work as a guide. For example, an item is defined as "prototypically immoral" by pre-set criterion ratings (i.e. an average of 75% of untimed participants regarded the item as immoral), the participant attributing immorality is encountering a "congruent" condition. This means that both the participant's automatic response and controlled response would result in the same determination (immoral action). The probability of identifying a congruent condition as immoral is quantified as the expression of control

probability, C, and the probability of automatic response occurring with the failure of control, A(1 – C):

$$P_{CONGRUENT} = C + A(1 - C)$$

In incongruent conditions, however, the participant attempts to make a judgment wherein their automatic response (to see intentionality or immorality) and controlled response (the item was rated only by 25% of untimed participants as immoral) are in conflict. In this situation, the likelihood of that the participant will identify the item as intentional or immoral is the probability of the expression of the automatic bias where there exists the failure of control, A(1 – C):

$$P_{INCONGRUENT} = A(1 - C)$$

Based on these assumptions, one can quantify separate estimates of controlled and automatic responding. Control estimates are defined as the difference between identifying the target in congruent (i.e. "correct") conditions and incongruent (i.e. "incorrect") conditions:

$$C = P_{CONGRUENT} - P_{INCONGRUENT}$$

Finally, with these conditions, one can solve for the automatic bias estimate as well:

$$A = P_{INCONGRUENT}/(1 - C)$$

According to this paradigm, therefore, controlled processes are defined as "those in which responses are successfully modulated by intentions" (p. 183) such that the individual is correct in processing the stimulus. Automatic processes, on the other hand, are processes that "operate regardless of whether they facilitate intentional performance or interfere with it" (p. 183) and therefore reflect a systematic preference to regard items as immoral or intentional. Each of these estimates are calculated using the same procedure, and categorized data according to the pre-set criterion in pilot testing with untimed control participants (Hester, Payne & Gray, 2017; Rosset, 2008).

Psychiatric symptoms

The Positive and Negative Syndrome Scale (PANSS, Kay et al., 1987) is an interview-based measure comprised of 30 items assessing for positive and negative symptoms of schizophrenia, as well as general psychopathology symptoms. These interviews were conducted and rated by experienced research assistants who were trained to adequate reliability (ICC > .80 with a gold standard rater). In the present study, we generated the five-factor solution subscales proposed by Bell and colleagues (1994): cognitive, emotional discomfort, hostility, positive, and negative symptoms, along with the specific item related to suspiciousness/persecution symptoms. Totals for each factor as calculated by Bell and colleagues (1994) are examined, as well as the specific score on the suspiciousness/persecution item separately.

Neurocognition

A verbal task of neurocognition was used in the discriminant validity analysis, given the paper-and-pencil vignette structure of the AIHQ. This 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).

Social cognitive skills

The final phase SCOPE battery was used to assess social cognitive skills in the present study. This battery consists of five measures. The Penn Emotion Recognition Task (ER-40; Kohler et al., 2003) is a 40-item assessment of emotion perception. Each item consists of a color photograph of a face that is expressing one of four states, happy, sad, angry, afraid, or neutral. Items represent a balanced presentation of gender, age, ethnicity, and intensity of emotion expressed. Participants are instructed to identify the emotion expressed as soon as possible after

seeing the face. Scores on this task range from 0 to 40, with each score representing the number of items answered correctly.

The Reading the Mind in the Eyes Task (Eyes; Baron-Cohen et al., 2001) is a 36-item assessment of theory of mind. Each item is presented as the eyes region of the face expressing a complex mental state. Participants are asked to determine what mental state is being depicted. Four options are presented with each photo. Scores range from 0 to 36, with each score representing the number of items answered correctly.

The Bell Lysaker Emotion Recognition Task (BLERT; Bryson, Bell, & Lysaker, 1997) assesses participants' ability to correctly affect in one actor presenting the same statement with varying emotional tones: happy, sad, afraid, disgusted, surprised, angry, or neutral. Each item is a ten-second clip of the face and shoulders of a male actor expressing the same statement with one of the affective tones. Scores range from 0 to 21 on this task, with higher scores indicating the number of correctly identified items (ranging from 0 to 21).

The Hinting Task (Corcoran, Mercer & Frith, 1995) involves participants interpreting ten brief written stories that require them to identify and make inferences involving others' mental states. Scores range from 0 to 20 on this task, with higher scores indicating better performance.

The Awareness of Social Inference Test - Social Inference: Minimal Subscale (TASIT, McDonald et al., 2004) consists of Yes/No questions related to four video-taped social vignettes requiring participants to infer individual motives which may contradict verbal communication (e.g., sarcasm or "white lies"). The TASIT is scored based on number of correct responses out of 60 possible**,** and includes subscales that distinguish between simple sarcasm (sarcastic phrases with a meaning that matches the utterance) and paradoxical sarcasm (phrases that imply the opposite of what they appear to express). While these both require the participant to represent the

internal state of the speaker to infer meaning, paradoxical items require a more complex judgment of meaning. Performance is indexed as total number correct.

Trait paranoia

The Persecution and Deservedness Scale (Melo et al., 2009) is a 10-item self-report scale designed to assess paranoia and perceived deservedness of persecution. Items describe traits or behaviors related to paranoia to which participants respond with a Likert scale response (scale 0 to 4) identifying the extent to which they identify with each item as well as a follow up item with the same scale identifying the extent to which they feel they deserve the reported persecution. This is designed to distinguish between bad me (depressive type) and poor me (non-affective psychosis type) paranoia in schizophrenia (Melo et al., 2009), though this distinction was not explored in the current study. Scores range from 0 to 40 with higher scores indicating higher levels of paranoia.

Trait hostility

The Personality Inventory for DSM-5 (PID-5, Krueger et al., 2012) is a 220-item self-report questionnaire evaluating potentially pathological personality dimensions related to DSM-5 disorders. Items consist of statements related to behaviors or personality dimensions and Likert scale (0 to 3) responses for participants. In the present study, participants were administered the ten items related to the Hostility Scale of the PID-5 (PID-5-HS). Total scores thus ranged from 0 to 30, with higher scores indicating greater hostility.

Observed hostility

The Observable Social Cognition: A Rating Scale (OSCARS; Healey et al., 2015) is a rating scale of the participant's performance in a number of arenas related to social cognition, including, for example, correctly understanding others' thoughts and intentions or jumping to

conclusions. There are eight items with accompanying Likert scale responses (1 = no evidence of difficulty to 7 = evidence of extreme difficulty). In the present study, we used the hostility item, which assesses whether the individual has difficulty "interpreting social interactions in a malevolent or hostile manner." For the current study, the informant-rated scale was used. Informants were identified by the participants and were high contact clinicians, family members, or close friends.

Role functioning

The Specific Levels of Functioning Scale (SLOF; Schneider & Struening, 1983) is a 31-item informant-rated measure of social functioning, community functioning, and effectiveness in activities of daily living. The present study examined the social acceptability subscale, which comprises the following items: regularly arguing with others, having physical fights with others, destroying property, physically abusing self, being fearful/crying/clinging, and taking property from others without permission. Ratings on the SLOF are made on a Likert scale as well (1 to 5, with higher scores indicating better functioning). Informants were the same as those selected for collection of the OSCARS. Given informant response rates, the sample for all analyses involving the SLOF (all conducted only in the schizophrenia group) were smaller than the full sample (n = 135).

Social skills

The Social Skills Performance Assessment (SSPA; Patterson et al, 2001) is an observer-rated assessment of social skill performance in two three-minute role-play conversations with a confederate. First, the participant is instructed to role-play a conversation with a new neighbor who has just moved to the area and second, a conversation with a landlord who had failed to fix a leak in the participant's house. The SSPA evaluates interest, speech fluency, clarity, focus,

affect, social appropriateness, submissiveness/persistence, negotiation ability, and overall effectiveness with scores summed and averaged into an overall score (ranging from 1 to 5). Average scores were calculated across both role-plays for the current study.

Procedure

Graduate-level staff, who had experience working with individuals with schizophrenia, conducted all interview-based measures. These research assistants were trained to reliability consistently across study sites. As mentioned, due to study protocol limitations with regard to requisite time for administration, the intentionality bias task was conducted at all research sites, but the immorality bias task was conducted only at one of three sites. Data collection took across two study visits, which were separated by an interval of 2 to 4 weeks. All variables reported here include only data from the initial study visit.

Data analytic plan

An overview of all aims, hypotheses, and planned statistical tests for main study aims can be found in Table 1, and for all exploratory aims in Table 2. Portions of analyses for hypotheses #1 through #3 utilize a multilevel model. Each task requires the participant to give a binary response to every single item. Although this kind of data has often been analyzed by collapsing participants' responses into a single value for each condition while conducting a within-subjects ANOVA, multilevel modeling conveys a few key advantages (Bauer & Curran, 2015). Multilevel models allow for the analysis of incomplete data, and they allow for the use a link function that treats the distribution of responses as binary, rather than normal. These models also account for within-subject variability (e.g., some participants might have a high baseline for immoral or intentional judgments, whereas other participants might have a low baseline).

The present analysis examined the effects of group, time pressure, and the inherent likelihood ratings (ILR) of each item. The baseline values are particularly important, given the fact that each item elicits a different baseline response (i.e. how likely it is that the individual will find each item to be the result of intentional or immoral action on the part of the hypothetical target). These baseline values were drawn from previous research on the immorality bias (Hester, Payne & Gray, 2017) and intentionality bias (Rosset et al., 2008). Thus, the present omnibus

Level 1 $Y_{ij} = \beta_{0j} + \beta_{1j}\text{percent} + r_{ij}$
Level 2 β_{0j} (intercept) $= \gamma_{00} + \gamma_{01}\text{time} + \gamma_{02}\text{group} + \gamma_{03}\text{timeXgroup} + u_{0j}$
β_{1j} (slope) $= \gamma_{10} + \gamma_{11}\text{time} + \gamma_{12}\text{group} + \gamma_{13}\text{timeXgroup} + u_{1j}$

Reduced model, not rewritten

$$Y = (\gamma_{00} + \gamma_{01}\text{time} + \gamma_{02}\text{group} + \gamma_{03}\text{timeXgroup} + u_{0j}) + (\gamma_{10} + \gamma_{11}\text{time} + \gamma_{12}\text{group} + \gamma_{13}\text{timeXgroup} + u_{1j})\text{percent} + r_{ij}$$

Reduced model, rewriting for effect and residuals

$$Y = [\gamma_{00} + \gamma_{01}\text{time} + \gamma_{02}\text{group} + \gamma_{03}\text{timeXgroup} + \gamma_{10}\text{Xpercent} + \gamma_{11}\text{timeXpercent} + \gamma_{12}\text{groupXpercent} + \gamma_{13}\text{timeXgroupXpercent}] + [u_{0j} + u_{1j}\text{Xpercent} + r_{ij}]$$

analysis is a multilevel model in which each item is nested within participant. The population expected response (i.e. "percent") is entered as a level 1 predictor, while the study manipulations (i.e. schizophrenia vs. control ["group"] and time pressure ["time"]) are entered as level 2 predictors, as shown below.

Each set of hypotheses and analyses are identical for the two examined measures here. For simplicity we will review hypotheses for each assessment in sequence, beginning with the intentionality bias task and then the immorality bias task.

CHAPTER 3: RESULTS

Demographics

All demographic analyses are reported in Table 3. The schizophrenia and control groups differed on one variable, education. The control group completed significantly more years of school than the schizophrenia group. With regard to study outcomes, there were significant differences between patients and controls in paranoia (higher in schizophrenia; PaDS), hostility (higher in schizophrenia; PID5-HS), social functioning (higher in controls; SSPA), and neurocognition estimate (higher in controls; WRAT).

Because only one of three of the study sites (Chapel Hill, NC) had participants complete the CIST, this subset was compared to the remainder of the full sample (at the Dallas, TX and Miami, FL sites) with regard to demographic variables and study outcomes. While there were no significant demographic differences between the subsets, they differed in a number of study outcome variables related to symptoms and functioning. Specifically, relative to the full sample, the CIST-subset was less paranoid, less hostile, less suspicious, experienced fewer interpersonal conflicts, and were higher functioning in work activities, independent living skills, social skills, and had better overall neurocognition.[3] Differences between the schizophrenia and control groups in hostility (PID5-HS) and neurocognition estimate (WRAT) were significantly smaller in the CIST-only subset. Overall, these differences reflect a tendency for all participants in the

[3]These analysis results refer to the main effect of subset, however this applies to patients only (because only patients completed these measures) for all the variables except for social skills and neurocognition, which both patients and control participants completed. With regard to social skills and neurocognition, there is a main effect of subset across schizophrenia and control groups.

CIST-only subset (both patients and controls) to be higher functioning and better educated and for specifically the schizophrenia sample participants in this subset, to be significantly less impaired, paranoid, and hostile. This might be the result of site differences, as the CIST-only subset was collected in a less urban, more affluent and higher educated area (Chapel Hill, NC) than the other two sites (Dallas, TX and Miami, FL), which were more urban and less affluent. This may be of particular interest given the relationship between schizophrenia incidence and urbanicity (Vassos et al., 2012).

Intentionality Bias Task (IBT) Analyses

Aim #1. Construct validity, group differences

After running the omnibus multilevel model, we examined the main effect of group to determine whether individuals with schizophrenia demonstrated an elevated bias toward intentionality. There was indeed a main effect of group, $F(1,7835) = 6.00$, $p < .014$. At the mean ILR rating, individuals in the schizophrenia group (M = 0.43, 95% CI [0.41, 0.46]) were more likely to identify items as intentional than the control group (M = 0.38, 95% CI [0.35, 0.41]).

Aim #2. Automatic and controlled processes

We tested the same full model as described above for a (1) a main effect of time pressure (expecting higher scores in high-time pressure) as well as (2) a two-way group by time pressure interaction (expecting greatest group differences in low time pressure condition), to examine our hypothesis that individuals with schizophrenia will be differentially affected by the time pressure manipulation relative to controls. There was no main effect of the time pressure manipulation, $F(1,7835) = 0.73$, $p = .40$. There was also no significant time pressure by group interaction, $F(1,7835) = 0.37$, $p = .54$, indicating that the rate of intentionality judgments of individuals in the schizophrenia sample were not differentially affected by the time pressure manipulation.

To examine the extent to which these differences were the result of automatic or controlled processing, we also examined parameters derived from the process dissociation procedure. First, with regard to the PDP-generated control estimate, there was a main effect of group, $F(1, 355) = 12.38$, $p < .001$, such that individuals with schizophrenia demonstrated lower control estimates than control participants across both conditions, and a main effect of time pressure, $F(1, 355) = 45.97$, $p < .001$, such that overall, participants demonstrated lower control in the fast condition than the slow condition. There was no time pressure by group interaction, $F(1, 355) = 0.80$, $p = .37$.

With regard to the automatic estimate, there was a main effect of group, $F(1, 360) = 6.72$, $p = .01$, such that individuals with schizophrenia demonstrated higher automatic bias than control participants across both conditions. There was no effect of time pressure condition, $F(1, 360) = 1.09$, $p = .30$, nor was there a group by time pressure interaction, $F(1, 360) = 0.05$, $p = .82$.

Taking these results together, it appears that individuals in general (regardless of diagnosis) make more judgments of intention under time pressure as a result of diminished recruitment of controlled processes. However, failure of control is not specific to individuals with schizophrenia. Instead, individuals with schizophrenia are more likely than controls to over-attribute intentionality as a result of an automatic bias that persists across higher and lower pressure conditions.

Aim #3. Psychometric analysis.

Convergent validity analyses were conducted with the expectation that the IBT bias scores would demonstrate significant relationships with hostility, positive, and suspiciousness symptoms as measured by the PANSS. This was not the case, as none of the IBT bias scores were significantly related to these symptom categories (Table 4).

Criterion validity analyses examined whether bias scores were related to related real-world outcomes, including interpersonal conflict, trait hostility, trait paranoia, and observed hostility. The IBT total score was significantly positively related to all of these outcomes; higher levels of bias toward intentionality related increased interpersonal conflict, hostility and paranoia. Pure bias and PDP-generated automatic estimates were not as consistently related to criterion outcomes; the pure bias estimate was only related to the hostility scale and trended toward significance in its relationship to the social acceptability scale of the SLOF. The PDP-generated automatic estimate was not related to any of these outcomes. Discriminant validity analyses showed that while the IBT total score and PDP-derived automatic bias scores were unrelated to the neurocognition estimate, the pure bias score had an unexpected negative relationship to the WRAT (Table 4).

Overall, these results indicate a number of patterns. First, it appears that the bias toward intentionality is not related to clinically-rated symptoms, however it does relate to functioning and separates from our estimate of neurocognition. Second, these results were most consistent when examining overall IBT total scores, rather than the pure bias score (which did no achieve discriminant validity from verbal intelligence) or the PDP-generated automatic estimate.

Exploratory Aim #1. Incremental validity

To examine the incremental validity of the IBT above and beyond the extant gold standard social cognitive skills battery in SCOPE, hierarchical linear regressions were performed to examine impact on significant criterion outcomes above and beyond the influence of the SCOPE battery. In all four hierarchical regressions, the gold standard social cognitive battery did not result in a significant model predicting any of the outcomes. When entered at step 2, the IBT total score significantly improved model fit with regard to interpersonal conflict ($\Delta R^2 = .03$, $p =$

.03), trait paranoia (ΔR^2 = .02, p = .03), and trait hostility (ΔR^2 = .02, p = .03), suggesting that the bias toward intentionality may provide unique information about criterion outcomes above and beyond the influence of skill-based measures. The full model predicting trait paranoia was significant overall (R^2 = .06, p = .04), while those for interpersonal conflict R^2 = .08, p = .09). and observed hostility approached significance (R^2 = .09, p = .08). These full models are reported in Table 6.

Exploratory Aim #2. Relationships beyond neurocognition

Both discernibility and PDP-derived control estimates were significantly related to neurocognition in the control group. In the schizophrenia group, however, only the PDP control estimate was significantly related to neurocognition (Table 5).

Exploratory Aim #3. Relationships to social functioning, social skills

The IBT was related to a number of general social and role functioning outcomes, as IBT total score and PDP-derived automatic bias estimate were significantly related to overall role functioning with greater bias being related to impaired functioning. The pure bias estimate was also negatively related to improved social skills, and all three IBT bias parameters were related to independent living skills with higher bias being related to poorer living skills. (Table 5).

Exploratory Aim #4. Relationships to non-psychosis psychopathology

With regard to negative and general symptoms, the IBT was significantly related to PANSS emotional discomfort symptoms; no other relationships were significant (Table 5).

Supplemental analysis #1. Differential response patterns by group.

In the process of examining the omnibus multi-level model, one additional effect emerged that is consistent with extant research in non-clinical samples (Hester, Payne & Gray, 2017). Specifically, there existed an ILR by group interaction, $F(1, 7835) = 20.82$, $p < .001$, such

that responses to items were most different between groups on items that have a low ILR rating. This pattern suggests that individuals with schizophrenia are most biased in items that are the least normatively regarded as intentional (Figure 1).

Categorizing Immoral Sentences Task (CIST) Analyses[4]

Aim #1. Construct validity, group differences

With regard to the CIST, there was no main effect of group, $F(1,2383) = 1.06$, $p = .31$, demonstrating that individuals with schizophrenia were no more likely to identify items as immoral than control participants.

Aim #2. Automatic and controlled processes

There also was no main effect of the time pressure manipulation, $F(1,2383) = 0.13$, $p = .72$, nor a time pressure by group interaction, $F(1,2383) = 0.02$, $p = .89$, indicating the time pressure interaction did not affect the responding of participants overall nor specifically schizophrenia participants. It is important to note that one central result in a non-clinical sample (Hester, Payne & Gray, 2017) was replicated as there did exist an interaction between ILR and time pressure condition combined across groups, $F(1,2383) = 5.54$, $p = .019$, such that all participants appeared to make more judgments of immorality in the fast time condition in the low-ILR (probably nonmoral) items (Figure 2).

With regard to the PDP-generated control estimates, there was no effect of group, $F(1, 92) = 0.39$, $p = .53$, or time pressure, $F(1, 92) = 1.58$, $p = .21$, however there was a time pressure by group interaction, $F(1, 92) = 5.17$, $p = .03$. Individuals with schizophrenia had higher control estimates ($M_{difference} = 0.04$, CI [-.17, .08]) in the fast condition but lower control in the slow condition ($M_{difference} = -0.14$, CI [-.24, -.04]). With regard to the automatic estimate, there was no

[4]The correlation between the total scores of the IBT and CIST are evaluated and found to be non-significant, $r = .13$, $p = .21$.

main effect of group, $F(1, 92) = 0.61$, $p = .43$, time pressure condition, $F(1, 92) = 0.01$, $p = .34$, nor a group by time pressure interaction, $F(1, 92) = 0.91$, $p = .34$.

Overall, individuals with schizophrenia do not seem to differ from controls with regard to the bias toward immorality, with one exception. When making such judgments, individuals with schizophrenia are indeed differentially affected by a time pressure manipulation; however, this failure of control did not result in a clear bias toward immorality. Generally speaking, these results do not support the presence of an elevated immorality bias in schizophrenia, nor one that is impacted by the recruitment of automatic and controlled cognitive processes, but rather, that individuals with schizophrenia are more likely to fail to recruit controlled cognitive processes in making such determinations relative to controls.

Aim #3. Psychometric analysis.

None of the convergent, criterion, or discriminant validity analyses revealed significant relationships (Table 4).

Exploratory Aim #1. Incremental validity

Because there were no significant relationships between the CIST and criterion outcomes, exploratory aim #1 was not completed for the CIST.

Exploratory Aim #2. Relationships beyond neurocognition

On the CIST, only the PDP control estimate was related to neurocognition in the control group. There was no relationship between either estimate of control and neurocognition in the schizophrenia group (Table 5).

Exploratory Aim #3. Relationships to social functioning, social skills

Only the pure bias estimate trended toward significance in its relationship to social skills as measured by the SSPA. None of the CIST parameters were related to any other measures of social and role functioning (Table 5).

Exploratory Aim #4. Relationships to non-psychosis psychopathology

None of the CIST parameters were related to any other symptom categories (Table 5).

Supplemental analysis #1. Differential response patterns by group.

Following the significant result of the analogous supplemental analysis on the IBT, we tested the same effect in CIST results, and found there was no ILR by group interaction, $F(1, 2383) = 0.22$, $p = 0.64$.

CHAPTER 4: DISCUSSION

The current study presents quite different conclusions across the two attribution biases examined, the intentionality bias and immorality bias. Results suggest that a bias toward intentionality may be a clinically relevant feature of psychotic disorders. This bias appears to be modestly related to interpersonal conflict, paranoia, trait hostility and observed hostility, though is not related to clinically-rated symptoms. Further, it provides variance predicting real-world interpersonal conflict outcomes above and beyond the influence of the gold-standard SCOPE battery. The immorality bias, on the other hand, does not appear to be related to any clinical or functional characteristics present in schizophrenia. Additionally, the present study provides new insights into methodological questions in the study of schizophrenia. These results lay the groundwork for continued research into the cognition involved in attributions of intention in schizophrenia, as well as implications for improvements using dual process paradigms in this population.

Intentionality Bias

Several previous findings on this bias were replicated and supported in the present study. First, individuals with schizophrenia sample appeared more likely to see others' actions as intentional relative to controls, corroborating previous work (Peyroux et al., 2014). Second, there appears to be a failure of control that takes place when the individual (regardless of diagnosis) has diminished access to cognitive resources (i.e. under time pressure), replicating initial work by Rosset (2008). Results deviated, however, in that this failure of control did not result in an

increased likelihood for study participants to identify items as intentional. The primary new contributions of the current study are that it demonstrates that this bias can be understood as (1) comprising both automatic and controlled components, (2) separate from clinically-rated positive, suspiciousness and hostility symptoms (differing from other attributional biases; Combs et al., 2007), and (3) related to interpersonal conflict outcomes above and beyond the influence of social cognitive skill.

With regard to automatic and controlled components of these judgments, individuals with schizophrenia showed *both* diminished control *and* increased bias according to estimates generated from process dissociation (Jacoby, 1991; Payne, 2001). This indicates that individuals with schizophrenia both are less apt to control their judgments *and* present with a bias to interpret intentional motives. We hypothesized that the increased rate of intentionality judgments in schizophrenia was a result of both an elevated bias as well as a diminished ability to regulate this bias under pressure; this was partially supported. The deficit in control and increase in bias in the schizophrenia group was not the result of differential response to pressure, as there were no interactions between time pressure and group membership. Individuals with schizophrenia do not appear to be differentially affected by the failure of control that occurs under time pressure. Instead, they appear more biased and less controlled regardless of pressure. Results here de-emphasize the centrality of a difficulty in regulating stress that leads to these biased social judgments, but rather a bias that is elevated regardless of context. Previous work has suggested that individuals with schizophrenia – and particularly those with persecutory delusions – not only struggle to identify and interpret the mental states of others, but also additionally overattribute intention to others (Moritz et al., 2011; Montag et al., 2011; Peyroux et al., 2014). The present

study provides support for the presence of this bias in schizophrenia and suggests that this bias is stable in the face of time pressure.

Second, this intentionality bias did not demonstrate convergent validity with hypothesized outcomes in the present study, including hostility, positive, and suspiciousness symptoms. This finding adds to a base of literature already rife with conflicting findings. Previous research suggests that detecting intention in clearly negative events appears related to persecutory symptoms (Buck et al., 2016b, 2017; Combs et al., 2007, 2009); in neutral events, however, this bias appears related to a diverse array of symptoms, including irritability and disorganization (Peyroux et al., 2014). While a clear model is unclear from the current results, results suggest that the intentionality bias captures a cognitive style non-redundant with clinically rated persecutory symptoms. In one exploratory analysis, the intentionality bias did appear related to general emotional distress symptoms (e.g. anxiety and dysphoria). This could suggest that an increased intentionality bias might be reflective of a presentation more consistent with depression, for example, if an individual regards others' intentional actions as stronger determinants of events in the world (rather than their own). These findings also raise questions as to whether this bias is specific to schizophrenia or more generally related to psychopathology. At present, firm conclusions about this relationship remain unclear, however, it is possible that a bias to regard all actions as intentional may increase unpleasant emotions, for example, fear or dislike of others when one experiences negative events.

Third, this bias showed significant relationships to a number of criterion outcomes, including interpersonal conflict, observed hostility, trait hostility, and trait paranoia, and it predicted these outcomes above and beyond the influence of all five measures comprising the gold-standard SCOPE social cognition battery (Pinkham et al., 2016). Further, exploratory

analyses demonstrated relationships with global living skills, social skills, and social functioning; these relationships – though small and not explicitly hypothesized – also are stronger than those demonstrated with extant measures of bias (Pinkham et al., 2016). These increases might be a result of the improvements to attributional style measurement tested here, including more items (Buck et al., 2017), a specific focus on determinations of intentionality (Peyroux et al., 2014), the minor stress-induction involved in timed responses (Moritz et al., 2011), or some combination of these and other factors.

Finally, a supplemental analysis provided additional potential explanations for the nature of the differences between individuals with schizophrenia and controls with regard to biased social reasoning. These analyses revealed an interaction between item characteristics (ILR) and group membership, such that less paradigmatically-intentional items produced the greatest group differences. This suggests that the differences between individuals with schizophrenia and controls in biased social judgments are greatest when items are paradigmatically accidental, not when they are ambiguous or paradigmatically intentional. This is consistent with recent suggestions that the abbreviation of attributional style measures to include only ambiguous events (as in the AIHQ) may be misguided, and that accidental items should be included as well (Buck et al., 2017). While in subclinical populations, ambiguous events elicit differential cognitive patterns, (Combs et al., 2007), in clinical studies, it appears that it's specifically the neutral or apparently accidental situations that may be interpreted differently in schizophrenia.

Overall, these results are encouraging and could provide additional insights into understanding aberrant social cognition in schizophrenia and the ways in which it might result in social dysfunction. This bias exists regardless of time pressure and is not exclusively an expression of a normative bias that emerges in the face of aberrant failure of control (Rosset,

2008). Instead, this bias is simply elevated in schizophrenia. Given its correlational nature, the current study cannot speak to a temporal model of the relationship between these biases, symptoms, and functioning. While additional prospective studies are needed, with further support, possibilities for explaining the impact on behavior and functioning of this bias are numerous. It could be the case that the tendency to over-attribute intentionality to others' actions results in more fearful, angry and dysphoric internal states when the individual experiences negative social events. This elevated bias could result in misunderstandings that increase disagreements, worsen communication, or lead to conflict. This might be exacerbated when these interpretations of intentionality occur in situations that typically regarded as accidental. Further, such dysfunction could compound as individuals avoid social interactions that they find threatening or confusing. Whether this bias is a cause, result, or a co-occurring feature of psychotic symptoms, it appears to identify a characteristic common to psychosis and its associated negative outcomes.

Immorality Bias

The existence of an elevated immorality bias in schizophrenia, on the other hand, did not receive support from the present study. There were no differences between groups or time pressure conditions, nor an interaction between them. It is notable, however, that the overall trend demonstrating a bias toward immorality (as in Hester, Payne & Gray, 2017) was indeed replicated, as items typically regarded as non-moral were more likely to be labeled as immoral when individuals were subject to time pressure. The present study provides additional support for the existence of a bias toward immorality under pressure overall; it provides no evidence, however, that this bias presents any differently in schizophrenia. The only significant group difference was a diminished control estimate among individuals with schizophrenia under time

pressure. As there were no differences in overall judgments or estimates of automatic bias, this appears to suggest that individuals with schizophrenia were simply less skilled at reading and responding to these items quickly, a pattern that (given the lack of bias) leads to random (and not biased) responding. Such a pattern would be expected among individuals with the cognitive impairments typically present in schizophrenia (Green et al., 2004; Twamley, Jeste & Bellack, 2003).

Importantly, there were two characteristics of the study that may have affected study results. First, the CIST analysis sample was considerably smaller than that of other study analyses. However, even with an anticipated larger sample, the data provided here do not provide support for medium or large effects. Second, and perhaps more importantly, there were a myriad of differences between the CIST-analysis subset and the overall sample with regard to symptoms, functioning, and other outcomes. These participants were all drawn from the research site that is less urban, more affluent and better educated. This sample demonstrated attenuated differences from controls with regard to neurocognition, functioning, and hostility, and thus they may not reflect a representative sample of individuals with schizophrenia.

However, while the diminished statistical power and characteristics of the sample require such interpretations to remain speculative, there still may be characteristics of this bias or measure that are flawed for clinical uses. It is possible that judgments of morality are simply not associated with the cognition specific to psychosis. As previous results have demonstrated that threat detection defined broadly is aberrant in schizophrenia (Green & Phillips, 2004), it may be the case that psychosis-related attributions are those that involve clear threat. Some immoral situations are removed from immediate danger but simply reflect bad behavior (e.g. “He lied to his sister”). Thus, it may also be the case that the structure of the task resulted in statements of

individuals' fixed abstract belief systems, rather than their responses to real-world interactions. In addition, previous studies (Combs et al., 2007; 2009) use second-person (rather than third-person) hypothetical vignettes, and it may be the case that these procedures more effectively elicit personal experiences. Future research might consider combining some of the improvements studied here with these strengths of extant measures.

Dual-process methodology in the study of attributions in psychosis

The present study examined not only two specific attribution biases, but also several methods aimed at examining cognitive biases in psychopathology in general, in particular, (1) time pressure manipulations, and (2) use of estimates of controlled and automatic responding, both those from individual-level logistic regression analyses and process dissociation procedure. Overall, each of these methods hold promise in the study of biases in psychotic disorders, but present results raise an array of questions related to them. First, the use of a time pressure manipulation appeared to be an effective method to examine changes in reliance on automatic and controlled processes that occur under stress. All individuals appeared to experience a failure of control under time pressure, but this did not lead to an increase in intentionality or immorality judgments. However, participants in the schizophrenia group presented with both an elevation in automatic bias and an amplified and consistent failure of control.

Previous methods used to examine attribution biases (AIHQ, Combs et al., 2007; ARAT, Fornells-Ambrojo & Garety, 2009) in schizophrenia have generally used total scores across items, and have not systematically examined agreement or disagreement from consensus-established item characteristics. The methods piloted here to examine this more specifically – *discernibility* (i.e. the agreement between each participant and normative responding), *pure bias* (i.e. the intercept of these models reflecting overall bias), *PDP-automatic* (i.e. preference for a

biased response based on dichotomous categories), and *PDP-control* (i.e. non-correspondence with pre-set dichotomous categories) – did not overall improve relationships to outcomes. Counter to expectations, results indicated that – in general – total scores appeared to be the best predictors of criterion outcomes in schizophrenia. On one hand, this might demonstrate that these new parameters are simply ineffective psychometric tools. On the other hand, this could indicate that total scores are the strongest predictors because aberrant attributions emerge from co-operation of automatic and controlled processes.

Limitations

Several limitations should be mentioned. Overall, one major limitation of the present study is drawing conclusions about social decision-making based on a controlled laboratory paradigm very removed from day-to-day life. The intentionality bias and immorality bias tasks are designed to elicit cognitive biases that might impact real-world emotions and behavior. However, they are laboratory-based tasks completed on computers. This method, though easily controlled and replicated, may present with decrements in ecological validity that deflate relationships between this paradigm and functional outcomes. This is a persisting question in this area, as there appears to exist a tradeoff involving internal and external validity between measurements that are distal or proximal to relevant functional outcomes (Brenner, Curbow & Legro, 1995).

There are additional limitations specific to methodology. In addition to concerns about sample size and site effects, the current study is also limited by a lack of prior results to inform correct specification of the reaction time tasks. With more time and resources, pilot testing of the paradigm with larger samples of both individuals with schizophrenia and non-clinical controls could help ensure that the task is designed properly for both populations. Second, though the

present study examines improvements in measurement of hostile attribution bias, an existing measure of this construct (i.e. the AIHQ) was not collected on this sample; instead, clinically-rated hostility, positive, and suspiciousness symptoms were used as convergent outcomes. Finally, some measurements of criterion outcomes used in the current study are proxies or brief versions. The WRAT, for instance, is a brief intelligence measure that, while having empirical support for this use (Weins et al., 1993), is still less psychometrically robust than the full-scale WAIS or MATRICS battery (Kern et al., 2009).

Directions for future research

For the intentionality bias, future research should both examine and test additional opportunities to improve the measure, as well as build an understanding of the model of its emergence and its potential clinical uses. In particular, given the results of supplemental analyses suggesting that judgments of intentionality differ most among paradigmatically-accidental items, the IBT could be strengthened by expanding the range of items to include more of these kinds of items. Further, future research should examine whether such cognitive biases precede and predict the emergence or worsening of symptoms. Previous research suggests that social cognitive bias tasks predict worsening of symptoms after a brief follow-up period (Buck et al., 2016b). It cannot be known based on the present study if or how a bias toward intentionality might play in the emergence of disorder related processes; if prospective studies provide support for this, the intentionality bias might provide another cognitive bias involved in the etiology of schizophrenia. Moreover, if these studies provide support for such a model, clinicians could practice encouraging individuals with psychosis to question initial impressions of others' intentions, as has been indicated in previous cognitive-behavioral treatment approaches (e.g. SCIT; Roberts, Penn & Combs, 2015).

With regard to the immorality bias task, the questions are more fundamental. Modifications should be made to the task and tested before continued use (including pilot testing at various millisecond increments), and improved editions should be tested on larger samples. On the other hand, future research could take the lessons of these null results and instead examine other biases that might better approximate the kinds of cognition relevant to psychosis. Some improvements might include changing vignettes to second person (Combs et al., 2007), including more items related describing non-threatening situations (Buck et al., 2017), as well as increasing emphasis on threat rather than abstract judgments.

Overall, the disparate, complex, yet promising results from the present study suggest that future research in the area of social cognitive biases in schizophrenia should be focused on a diverse array of biases, methodologies, and models of their emergence. As has been the case with previous social cognitive biases in psychosis (e.g. trustworthiness bias, hostile attribution bias), each construct appears to have its own complex profile and relationship to outcomes in this population. The present study makes a first-pass effort at applying a number of new methods to the study of social cognitive biases in schizophrenia, and as with any new application, such an endeavor provides significant promise as well as a number of limitations to address and questions to answer.

APPENDIX 1: FIGURES

Figure 1. Comparing responses (in logistic regression) in the schizophrenia and control groups graphed as a function of inherent likelihood of responding for the Intentionality Bias Task (IBT) combined across time condition.

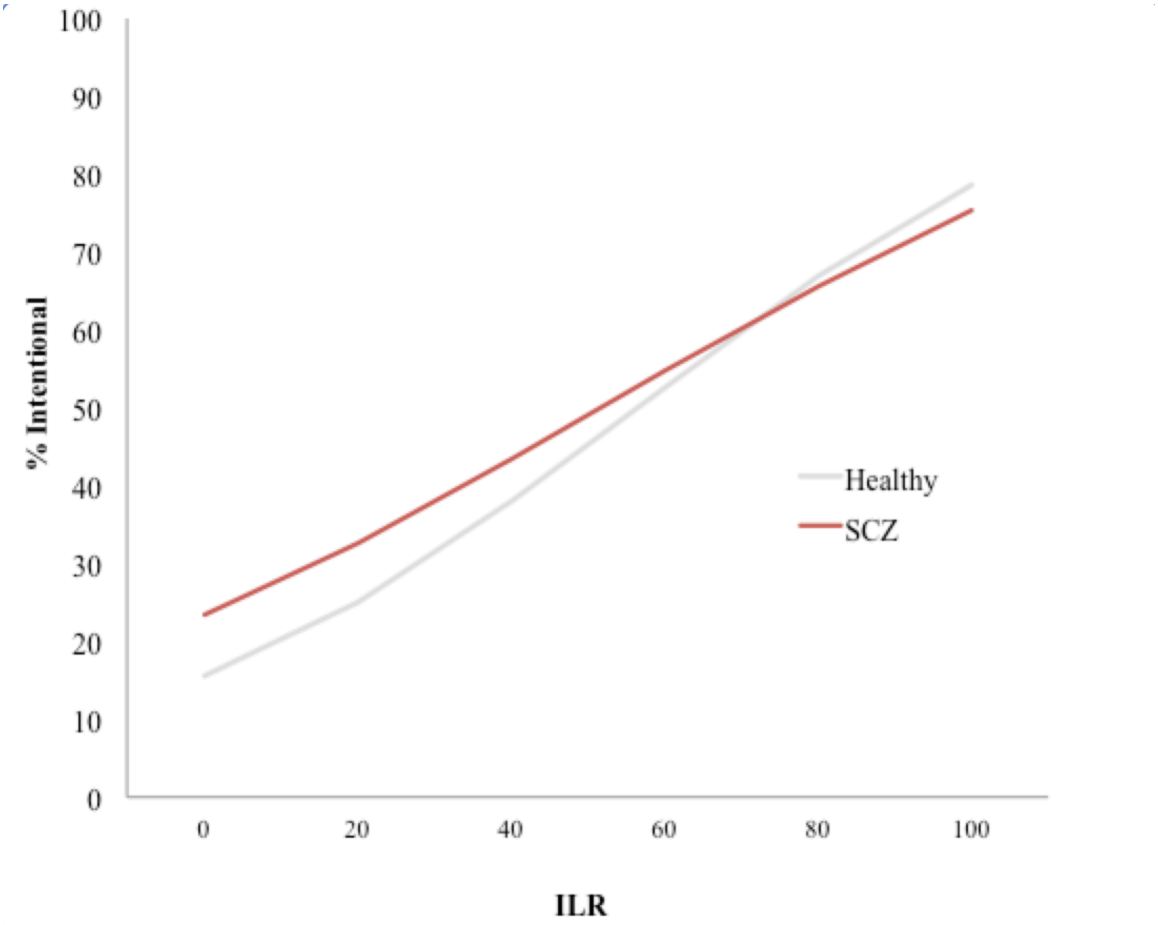
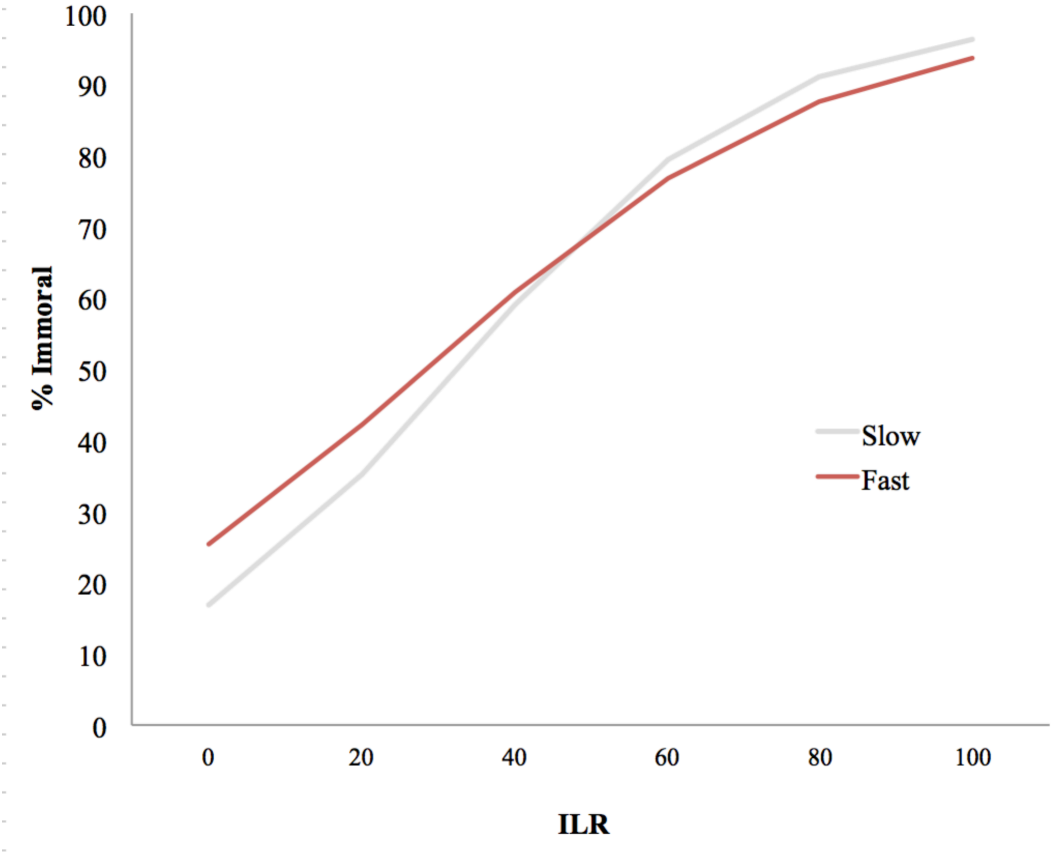


Figure 2. Comparing responses (in logistic regression) in the fast and slow conditions, graphed as a function of inherent likelihood of responding for the immorality bias (CIST) combined across groups.



APPENDIX 2: TABLES

Table 1. An overview of the main aims, hypotheses, and related statistical tests in the proposed study.

Aim	Hypothesis	Test	Outcome	Predictor	Prediction
#1. Group differences.	Individuals with schizophrenia will make more attributions of intentionality and evil on both measures.	Multi-level model	Intentionality responses Immorality responses	Group	SCZ > CTRL
#2. Automatic and controlled processes.	Individuals with schizophrenia will demonstrate elevated biases, as well as deficits in the ability to make controlled judgments across low and high time pressure conditions.	Multi-level model	Intentionality responses Immorality responses	Time pressure Group x Time Interaction	Fast > Slow Interaction (↑ differences between groups in slow condition)
		Mixed-model ANOVA.	Automatic estimates (PDP)	Group	SCZ > CTRL
			Controlled estimates (PDP)	Group Time pressure Group x Time Interaction	CTRL > SCZ Slow > Fast Interaction (↑ differences between groups in slow condition)
#3. Psychometric validation	The immorality and intentionality biases will demonstrate sufficient psychometric characteristic as measures of hostile attributions in schizophrenia with regard to:	Multi-level model	Intentionality responses Immorality responses	Group	SCZ > CTRL
	Convergent validity	Pearson correlations	Total score Pure bias estimate Automatic estimate (PDP)	Hostility symptoms Positive symptoms Susp./Pers. symptoms	All positive (with the exception of Social Acceptability, negative)
	Criterion validity			Social acceptability Trait hostility Trait paranoia Observed hostility	
	Discriminant			Neurocognition	Non-significant

Table 2. An overview of the exploratory aims, hypotheses, and related statistical tests in the proposed study.

Aim		Test	Outcome	Predictor
#1. Incremental validity of biases above social cognitive skill.	Examine whether these new measures add variance in predicting criterion outcomes above and beyond existing measures of social cognitive skill.	Hierarchical linear regression	Pure bias (collapsed across time condition) After Step 1 of all SCOPE social cognitive skill measures	(Those significant in aim #3) Social acceptability Trait hostility Trait paranoia Aggressive behavior Observed hostility
#2. Control and neurocognition.	Examine relationship of discernibility and control estimate to neurocognition.	Pearson correlation	Discernibility Control estimate (PDP)	Neurocognition
#3. Bias measures and functioning.	Examine relationship of immorality and intentionality biases to measures of functioning.	Pearson correlation	Total score Pure bias Automatic estimate (PDP)	Role functioning (total and subscales) Social skills
#4. Bias measures and general distress.	Examine relationship of immorality and intentionality biases to general non-psychosis related psychopathology symptoms.	Pearson correlation	Total score Pure bias Automatic estimate (PDP)	Cognitive symptoms Emotional discomfort symptoms Negative symptoms

Table 3. Participant demographics and tests for differences between the schizophrenia and non-clinical control samples.

	Group				
	SCZ full *(IBT only)* (*n* = 217)	**Control full** *(IBT only)* (n =154)	**SCZ subset** *(IBT/CIST)* (*n* = 42)	**Control subset** *(IBT/CIST)* (n = 58)	Effect
Demographics					
Age	41.71 (11.67)	41.95 (12.42)	39.82 (12.04)	41.07 (12.96)	
Education (years)	13.04 (2.49)	14.19 (1.91)	14.38 (2.25)	15.15 (1.66)	**+++
Gender					
Male	141 (64.98%)	97 (62.99%)	30 (71.43%)	39 (67.24%)	
Female	76 (35.02%)	57 (37.01%)	12 (28.58%)	19 (32.76%)	
Race					**&**
White	115 (53.00%)	80 (51.95%)	33 (78.57%)	32 (55.17%)	
Black	87 (40.10%)	62 (40.26%)	7 (16.67%)	23 (39.66%)	
Am Ind / PI	3 (1.39%)	0 (0.00%)	1 (2.38%)	0 (0.00%)	
Asian	5 (2.30%)	4 (2.60%)	0 (0.00%)	1 (1.72%)	
Other	7 (3.22%)	8 (5.19%)	1 (2.38%)	2 (3.44%)	
Symptoms					
PANSS Cognitive	2.06 (0.69)	--	2.09 (0.58)	--	
PANSS Hostility	1.43 (0.61)	--	1.43 (0.48)	--	
PANSS Susp.	2.80 (1.42)	--	2.40 (1.23)	--	+
PANSS Emotional Distress	2.66 (1.12)	--	2.60 (0.99)	--	
PANSS Negative	1.74 (0.75)	--	1.92 (0.75)	--	**&**
PANSS Positive	2.52 (0.91)	--	2.38 (0.80)	--	
Criterion outcomes					
OSCARS Hostility Item	2.77 (1.76)	--	2.09 (1.10)	--	++
PADS Persecution (avg)	1.52 (1.07)	0.60 (0.68)	1.33 (0.99)	0.68 (0.72)	***
PID-5 Hostility Scale	9.60 (6.99)	5.92 (4.87)	7.69 (5.42)	6.64 (5.55)	***^
SLOF Social Acceptability	4.43 (0.58)	--	4.71 (0.30)	--	++
General functional outcomes					
SLOF Total	4.12 (0.64)	--	4.37 (0.52)	--	++
SLOF Work	3.72 (0.98)	--	4.18 (0.79)	--	++
SLOF Activities	4.54 (0.93)	--	4.69 (0.69)	--	
SSPA Total	4.14 (0.54)	4.54 (0.38)	4.51 (0.40)	4.36 (0.44)	***++
UPSA-B Total	70.80 (14.04)	--	77.65 (11.41)	--	+++
Neurocognition estimate					
WRAT	94.87 (14.62)	101.11 (11.48)	102.69 (11.48)	102.29 (10.12)	*+++^

Main effect of SCZ vs. Control across the full sample ($p < .05$, **$p < .01$, ***$p<.001$)
+Main effect of subset vs. full across clinical or control group (+$p < .05$, ++$p < .01$, +++$p<.001$)
^Interaction of subset/full by group (^$p < .05$, ^^$p < .01$, ^^^$p<.001$)
& The differences in racial breakdown between the groups approached significance in the CIST-only subset, $\chi^2 = 8.54$, sig. = .074. Additionally, the difference between the subsets was significant with regard to suspiciousness $F(1, 215) = 4.103$, $p = .04$, and approached significance with regard to negative symptoms, $F(1, 215) = 3.05$, $p < .08$.
Note: A subset of participants completed both the IBT and CIST. These participants' demographics and means for study variables are reported here in the "(full)" columns. The demographic and study variable means for the subset are reported for each group above in the "(CIST)" columns. PANSS, PADS, SLOF and SSPA scores are reported as item averages.

Table 4. Psychometric examination of the intentionality and immorality bias tasks (Hypothesis #3). All analyses are Pearson correlations.

	Intentionality Bias (IBT)			**Immorality Bias (CIST)**		
	Total Score	Pure Bias	PDP Automatic	Total Score	Pure Bias	PDP Automatic
Hypothesis #3 – Psychometric validation						
Convergent validity – Hostility, positive, and suspiciousness symptoms						
PANSS Hostility	.02	.11	.03	-.12	-.18	-.09
PANSS Positive	-.05	-.05	-.09	.14	.02	.12
PANSS Suspiciousness	.03	.04	.00	.05	-.24	.04
Criterion validity – Social functioning						
SLOF – Social Accept.	**-.21***	**-.16^**	-.13	-.17	-.03	-.19
PID5 – Hostility Scale	**.15***	**.15***	.09	.24	-.16	.18
PADS – Paranoia	**.13***	-.04	.06	.08	.03	.07
OSCARS – Hostility	**.17***	.09	.12	-.11	-.09	-.20
Discriminant validity – Intellectual functioning						
WRAT	-.06	**-.38*****	-.10	-.03	-.06	-.05

Note: Total participant totals for outcome measures differed from study totals, including the SLOF (n = 135) and OSCARS (n = 133) given response rates of informants.
Total scores of the intentionality and immorality bias tasks are uncorrelated, r = -.04, p = .79.
***p < .001, **p < .01, *p < .05, ^p < .10

Table 5. Additional psychometric examination of the intentionality and immorality bias tasks (Exploratory Aims #2 through #4).

	Intentionality Bias (IBT)		**Immorality Bias (CIST)**	
	Discernibility Score	PDP Control	Discernibility Score	PDF Control
Exploratory Aim #2 – Intellectual functioning and control				
WRAT (CTRL)	**.19***	**.30*****	.05	**.32***
WRAT (SCZ)	.02	**.36*****	.14	.25

	Intentionality Bias (IBT)			**Immorality Bias (CIST)**		
	Total Score	Pure Bias	PDP Automatic	Total Score	Pure Bias	PDP Automatic
Exploratory Aim #3 – General functioning						
SLOF – Total	**-.17***	-.14	**-.16***	-.17	-.22	-.09
SLOF – Work performance	-.11	-.14	-.07	-.14	-.09	-.05
SLOF – Activities	-.11	-.08	**-.17^**	-.06	-.20	-.07
SSPA Total	-.05	**-.20****	-.10	.08	**.27^**	.10
UPSA Total	**-.15***	**-.25*****	**-.18***	-.03	-.25	.03
Exploratory Aim #4 – Negative and general symptoms						
PANSS Cognitive	-.06	.06	.00	-.17	-.16	-.13
PANSS Emotional Distress	**.20****	.08	.09	.08	.08	.11
PANSS Negative	-.07	-.08	-.09	-.01	-.17	.02

Note: Total participant totals for outcome measures differed from study totals, including the SLOF (n = 135) and OSCARS (n = 133) given response rates of informants. Additionally, six participants did not complete the SSPA (n = 210), and eight did not complete the UPSA (n = 208). ***p < .001, **p < .01, *p < .05, ^p < .10

Table 6. Incremental validity analyses examining partial 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 item totals (all in the schizophrenia sample only).

	Interpersonal Conflict			**Trait Hostility**			**Trait Paranoia**			**Hostile Behavior**		
	SLOF Social Acceptability			**PID5 Hostility Scale**			**PADS Total**			**OSCARS Hostility Item**		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Step 1 – SCOPE Social Cognition Skills												
BLERT	***0.04***	***0.02***	***0.22^***	***-.0.20***	***0.19***	***-0.11^***	***-0.07***	***0.03***	***-0.27****	***-.17***	***0.06***	***-0.34****
ER-40	*-0.02*	*0.02*	*-0.13*	*0.05*	*0.15*	*0.03*	*0.03*	*0.02*	*0.12*	*0.03*	*0.05*	*0.06*
Eyes Test	*0.01*	*0.01*	*0.10*	*-0.06*	*0.12*	*-0.05*	*0.02*	*0.02*	*0.10*	*0.03*	*0.04*	*0.10*
Hinting Test	*0.01*	*0.02*	*0.05*	*-0.03*	*0.14*	*-0.01*	*0.03*	*0.02*	*0.10*	*-0.05*	*0.05*	*-0.09*
TASIT Total	*-0.01*	*0.01*	*-0.15*	*0.02*	*0.08*	*0.02*	*0.01*	*0.01*	*0.06*	*0.04*	*0.03*	*0.14*
	(R^2 = .05, p = .27)			(R^2 = .02, p = .68)			(R^2 = .04, p = .13)			(R^2 = .07, p = .12)		
Step 2 – IBT Total												
BLERT	0.03	0.02	0.20	-0.22	0.18	-0.13	**-0.07**	**0.03**	**-0.28****	**-0.16**	**0.06**	**-0.32***
ER-40	-.0.02	0.02	-0.14	0.10	0.15	0.06	0.03	0.02	0.15	0.03	0.05	0.07
Eyes Test	0.01	0.01	0.10	-0.04	0.12	-0.03	0.02	0.02	0.11	0.03	0.04	0.10
Hinting Test	0.01	0.02	0.04	-0.01	0.14	-0.01	0.03	0.02	0.10	-0.04	0.05	-0.08
TASIT Total	-0.01	0.01	-0.10	-0.00	0.08	-0.00	0.01	0.01	0.04	0.03	0.03	0.10
Intentionality Bias Total	***-0.03***	***0.01***	***-0.19****	***0.26***	***0.12***	***0.15****	***0.04***	***0.02***	***0.15****	***0.07***	***0.04***	***0.14***
	(ΔR^2 = .03, p = .03)*			**(ΔR^2 = .02, p = .03)***			**(ΔR^2 = .02, p = .03)***			(ΔR^2 = .02, p =.11)		
Full Model R^2	**(R^2 = .08, p = .09)^**			(R^2 = .04, p = .26)			**(R^2 = .06, p = .04)***			**(R^2 = .09, p = .08)^**		

Note: Sample sizes for these analyses are consistent (n = 217), with some exception, including the SLOF (n = 135) and OSCARS (n = 133) given response rates of informants.
***p < .001, **p < .01, *p < .05, ^p < .10

REFERENCES

Aakre, J. M., Seghers, J. P., St-Hilaire, A., & Docherty, N. (2009). Attributional style in delusional patients: a comparison of remitted paranoid, remitted nonparanoid, and current paranoid patients with nonpsychiatric controls. *Schizophrenia Bulletin 35*(5), 994-1002.

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology*, *87*(1), 49-74.

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. *Schizophrenia Research 93*(1), 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. *Schizophrenia Research 118*(1), 54-61.

Baron‐Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high‐functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241-251

Bauer, D.J. & Curran, P.J. (2015). *Multilevel Modeling.* Department of Psychology, L.L Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill.

Bell, M. D., Lysaker, P. H., Beam-Goulet, J. L., Milstein, R. M., & Lindenmayer, J. P. (1994). Five-component model of schizophrenia: assessing the factorial invariance of the positive and negative syndrome scale. *Psychiatry Research*, *52*(3), 295-303

Bellack, A. S., Green, M. F., Cook, J. A., Fenton, W., Harvey, P. D., Heaton, R. K., Laughren, T., Leon, A.C., Mayo, D.J., Patrick, D.L., Patterson, T.L., Rose, A., Stover, E. & Wykes, T. (2007). Assessment of community functioning in people with schizophrenia and other severe mental illnesses: a white paper based on an NIMH-sponsored workshop. *Schizophrenia Bulletin 33*(3), 805-822.

Bentall, R. P., & Kaney, S. (2005). Attributional lability in depression and paranoia. *British Journal of Clinical Psychology*, *44*(4), 475-488.

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

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: meta-analysis. *Schizophrenia Research 109*(1), 1-9.

Brakoulias, V., Langdon, R., Sloss, G., Coltheart, M., Meares, R., & Harris, A. (2008). Delusions and reasoning: a study involving cognitive behavioural therapy. *Cognitive Neuropsychiatry*, *13*(2), 148-165.

Brenner, M. H., Curbow, B., & Legro, M. W. (1995). The proximal-distal continuum of multiple health outcome measures: the case of cataract surgery. *Medical Care 33*(4), AS236-AS244.

Browne, J., Penn, D.L., Raykov, T., Pinkham, A.E., Kelsven, S., Buck, B., Harvey, P.D. (2016). Social Cognition in Schizophrenia: Factor Structure of Emotion Processing and Theory of Mind. *Psychiatry Research 242*, 150-156.

Bryson, G., Bell, M., & Lysaker, P. (1997). Affect recognition in schizophrenia: a function of global impairment or a specific cognitive deficit. *Psychiatry Research*, *71*(2), 105-113.

Buck, B., Healey, K., Gagen, E., Penn, D.L. (2016a). Social cognition in schizophrenia: Factor structure, clinical and functional correlates. *Journal of Mental Health 25*(4), 330-337.

Buck, B., Pinkham, A., Harvey, P.D., Penn, D.L. (2016b). Revisiting the validity of measures of social cognitive biases in schizophrenia: Results from the Social Cognition Psychometric Evaluation (SCOPE) study. *British Journal of Clinical Psychology 55*(4), 441-454. *Journal of Clinical Psychology.*

Buck, B., Iwanski, C., Healey, K.M., Green, M.F., Horan, W.P., Kern, R., Lee, J., Marder, S., Reise, S.P. Penn, D.L. (2017). Improving measurement of attributional style in schizophrenia; A psychometric evaluation of the Ambiguous Intentions Hostility Questionnaire (AIHQ). *Journal of Psychiatric Research 89,* 48-54.

Candido, C.L., & Romney, D.M. (1990). Attributional style in paranoid vs. depressed patients. *British Journal of Medical Psychology 63*(4), 355-363.

Chang, H., Lee, S.K., Kim, K.R., Lee, S.Y., Park, J.Y., Kim, E.J., Kang, J. I., Lee, E. & An, S.K. (2009). Development of Korean version of the ambiguous intentions Hostility Questionnaire (K-AIHQ). *Journal of the Korean Neuropsychiatric Association 48*(1), 29-35.

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical psychology review*, *30*(2), 203-216

Combs, D. R., Adams, S. D., Penn, D. L., Roberts, D., Tiegreen, J., & Stem, P. (2007a). Social Cognition and Interaction Training (SCIT) for inpatients with schizophrenia spectrum disorders: preliminary findings. *Schizophrenia Research*, *91*(1), 112-116.

Combs, D. R., Finn, J. A., Wohlfahrt, W., Penn, D. L., & Basso, M. R. (2013). Social cognition and social functioning in nonclinical paranoia. *Cognitive Neuropsychiatry*, *18*(6), 531-548.

Combs, D. R., Penn, D. L., Chadwick, P., Trower, P., Michael, C. O., & Basso, M. R. (2007b). Subtypes of paranoia in a nonclinical sample. *Cognitive Neuropsychiatry 12*(6), 537-553.

Combs, D. R., Penn, D. L., Michael, C. O., Basso, M. R., Wiedeman, R., Siebenmorgan, M., Tiegreen, J. & Chapman, D. (2009). Perceptions of hostility by persons with and without persecutory delusions. *Cognitive Neuropsychiatry 14*(1), 30-52.

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry 12*(2), 128-143.

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

Couture, S. M., Penn, D. L., & Roberts, D. L. (2006). The functional significance of social cognition in schizophrenia: a review. *Schizophrenia Bulletin 32*(S1), 44-63.

Couture, S. M., Penn, D. L., Losh, M., Adolphs, R., Hurley, R., & Piven, J. (2010). Comparison of social cognitive functioning in schizophrenia and high functioning autism: more convergence than divergence. *Psychological Medicine*, *40*(04), 569-579.

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

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin 115*(1), 74-101.

DelVyder, J. E., Ben-David, S., Kimhy, D., & Corcoran, C. M. (2013). Attributional style among youth at clinical risk for psychosis. *Early Intervention in Psychiatry 7*(1), 84-88.

Diez-Alegría, C., Vázquez, C., Nieto-Moreno, M., Valiente, C., & Fuentenebro, F. (2006). Personalizing and externalizing biases in deluded and depressed patients: Are attributional biases a stable and specific characteristic of delusions? *British Journal of Clinical Psychology 45*(4), 531-544.

Donohoe, G., Spoletini, I., McGlade, N., Behan, C., Hayden, J., O'Donoghue, T., Peel, R., Haq, F., Walker, C., O'Callaghan, E., Spalletta, G., Goll, M., Corvin, A. & Corvin, A. (2008). Are relational style and neuropsychological performance predictors of social attributions in chronic schizophrenia? *Psychiatry Research 161*(1), 19-27.

Eckman, T.A., Wirshing, W.C., Marder, S.R., Liberman, R.P., Johnston-Cronk, K., Zimmermann, K., & Mintz, J. (1992). Technique for training schizophrenic patients in illness self-management: A controlled trial. *American Journal of Psychiatry 149*(11), 1549-1555.

Elnakeeb, M., Abdel-Dayem, S., Gaafar, M., & Mavundla, T. R. (2010). Attributional style of Egyptians with schizophrenia. *International Journal of Mental Health Nursing 19*(6), 445-456.

Falloon, I. R., Boyd, J. L., McGill, C. W., Razani, J., Moss, H. B., & Gilderman, A. M. (1982). Family management in the prevention of exacerbations of schizophrenia: a controlled study. *New England Journal of Medicine 306*(24), 1437-1440.

Fett, A. K. J., Viechtbauer, W., Dominguez, M. D. G., Penn, D. L., van Os, J., & Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional

outcomes in schizophrenia: a meta-analysis. *Neuroscience & Biobehavioral Reviews 35*(3), 573-588.

Fitch, K., Bernstein, S. J., Aguilar, M. D., Burnand, B., & LaCalle, J. R. (2001).*The RAND/UCLA appropriateness method user's manual* (No. RAND/MR-1269-DG-XII/RE). Rand Corporation: Santa Monica, California.

Fornells-Ambrojo, M., & Garety, P. A. (2009). Attributional biases in paranoia: the development and validation of the Achievement and Relationships Attributions Task (ARAT). *Cognitive Neuropsychiatry*, *14*(2), 87-109.

Fraguas, D., Mena, A., Franco, C., Martín-Blas, M. M., Nugent, K., & Rodríguez-Solano, J. J. (2008). Attributional style, symptomatology and awareness of illness in schizophrenia. *Psychiatry Research 158*(3), 316-323.

Gold, J. M. (2004). Cognitive deficits as treatment targets in schizophrenia. *Schizophrenia Research 72*(1), 21-28.

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*(4), 1600-1615.

Green, M. F., & Leitman, D. I. (2008). Social cognition in schizophrenia. *Schizophrenia Bulletin 34*(4), 670-672.

Green, M. F., Nuechterlein, K. H., Gold, J. M., Barch, D. M., Cohen, J., Essock, S., Fenton, W.S., Frese, F., Goldberg, T.E., Heaton, R.K., Keefe, R.S.E., Kern, R.S., Kraemer, H., Stover, E., Weinberger, D.R., Zalcman, S. & Marder, S. R. (2004). Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. *Biological Psychiatry 56*(5), 301-307.

Green, M. F., Lee, J., & Ochsner, K. N. (2013). Adapting social neuroscience measures for schizophrenia clinical trials, part 1: Ferrying paradigms across perilous waters. *Schizophrenia Bulletin 39*(6), 1192-1200.

Green, M. F., Kern, R. S., Braff, D. L., & Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the" right stuff"?. *Schizophrenia Bulletin*, *26*(1), 119-136.

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. *Schizophrenia Bulletin 34*(6), 1211-1220.

Hall, J., Harris, J. M., Sprengelmeyer, R., Sprengelmeyer, A., Young, A. W., Santos, I. M., Johnstone, E.C., & Lawrie, S. M. (2004). Social cognition and face processing in schizophrenia. *The British Journal of Psychiatry*, *185*(2), 169-170.

Hasson-Ohayon, I., Mashiach-Eizenberg, M., Avidan, M., Roberts, D. L., & Roe, D. (2014). Social Cognition and Interaction Training: Preliminary Results of an RCT in a Community Setting in Israel. *Psychiatric Services 65*(4). 555-558.

Heimberg, R. G., Klosko, J. S., Dodge, C. S., Shadick, R., Becker, R. E., & Barlow, D. H. (1989). Anxiety disorders, depression, and attributional style: A further test of the specificity of depressive attributions. *Cognitive Therapy and Research*, *13*(1), 21-36.

Healey, K. M., Combs, D. R., Gibson, C. M., Keefe, R. S., Roberts, D. L., & Penn, D. L. (2015). Observable Social Cognition–A Rating Scale: an interview-based assessment for schizophrenia. *Cognitive Neuropsychiatry*, *20*(3), 198-221.

Hester, N., Payne, B. K., & Gray, K. (2017). *The Immorality Bias: Why "John Flurbed Mary" Seems Wrong,* Unpublished manuscript, Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill.

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

Hooker, C. I., Tully, L. M., Verosky, S. C., Fisher, M., Holland, C., & Vinogradov, S. (2011). Can I trust you? Negative affective priming influences social judgments in schizophrenia. *Journal of Abnormal Psychology*, *120*(1), 98-w.

Holthausen, E. A., Wiersma, D., Cahn, W., Kahn, R. S., Dingemans, P. M., Schene, A. H., & van den Bosch, R. J. (2007). Predictive value of cognition for different domains of outcome in recent-onset schizophrenia. *Psychiatry Research 149*(1), 71-80.

Horan, W. P., Green, M. F., DeGroot, M., Fiske, A., Hellemann, G., Kee, K., ... & Nuechterlein, K. H. (2012). Social cognition in schizophrenia, part 2: 12-month stability and prediction of functional outcome in first-episode patients. *Schizophrenia Bulleti*n *38*(4), 865-872.

Horan, W. P., Kern, R. S., Shokat-Fadai, K., Sergi, M. J., Wynn, J. K., & Green, M. F. (2009). Social cognitive skills training in schizophrenia: an initial efficacy study of stabilized outpatients. *Schizophrenia Research 107*(1), 47-54.

Horan, W. P., Kern, R. S., Tripp, C., Hellemann, G., Wynn, J. K., Bell, M., ... & Green, M. F. (2011). Efficacy and specificity of social cognitive skills training for outpatients with psychotic disorders. *Journal of Psychiatric Research 45*(8), 1113-1122.

Humphreys, L. & Barrowclough, C. (2006). Attributional style, defensive functioning and persecutory delusions: Symptom-specific or general coping strategy?. *British Journal of Clinical Psychology 45*(2), 231-246.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541.

Jeon, I. H., Kim, K. R., Kim, H. H., Park, J. Y., Lee, M., Jo, H. H., Koo, S.J., Jeong, Y.J., Song, Y.Y., Kang, J.I., Lee, S.Y., Lee, E. & An, S. K. (2013). Attributional Style in Healthy Persons: Its Association with Theory of Mind Skills. *Psychiatry Investigation 10*(1), 34-40.

Jolley, S., Garety, P., Bebbington, P., Dunn, G., Freeman, D., Kuipers, E., Fowler, D. & Hemsley, D. (2006). Attributional style in psychosis—the role of affect and belief type. *Behaviour Research and Therapy 44*(11), 1597-1607.

Kanie, A., Hagiya, K., Ashida, S., Pu, S., Kaneko, K., Mogami, T., Oshima, S., Motoya, M., Niwa, S., Inagaki, A., Ikebuchi, E., Kikuchi, A., Yamasaki, S., Iwata, K., Roberts, D.L. & Nakagome, K. (2014). New instrument for measuring multiple domains of social cognition: Construct validity of the Social Cognition Screening Questionnaire (Japanese version). *Psychiatry and Clinical Neurosciences 68*(9), 701-711.

Kay, S.R., Fiszbein, A., & Opler, L. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia Bulletin, 13, 261-274.

Kern, R. S., Gold, J. M., Dickinson, D., Green, M. F., Nuechterlein, K. H., Baade, L. E., Keefe, R.E., Mesholam-Gately, R.I., Seidman, L.J., Lee. C., Sugar, C. A., & Marder, S. (2011). The MCCB impairment profile for schizophrenia outpatients: results from the MATRICS psychometric and standardization study. *Schizophrenia Research*, *126*(1), 124-131.

Kim, N. W., Song, Y. Y., Park, J. Y., Baek, S. Y., Kang, J. I., Lee, E., & An, S. K. (2014). Emotional dysregulation, attributional bias, neurocognitive impairment in individuals at ultra-high risk for psychosis and with schizophrenia: its association with paranoia. *Korean Journal of Schizophrenia Research*, *17*(2), 63-71.

Kinderman, P., & Bentall, R. P. (1996). A new measure of causal locus: the internal, personal and situational attributions questionnaire. *Personality And Individual Differences 20*(2), 261-264.

Kinderman, P., & Bentall, R. P. (1997b). Causal attributions in paranoia and depression: internal, personal, and situational attributions for negative events. *Journal Of Abnormal Psychology 106*(2), 341-345.

Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology 16*, 309–324.

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., & Gur, R. C. (2003). Facial emotion recognition in schizophrenia: intensity effects and error pattern. *American Journal of Psychiatry*, *160*(10), 1768-1774.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2009). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophrenia Bulletin 36*(5), 1009-1019.

Krstev, H., Jackson, H., & Maude, D. (1999). An investigation of attributional style in first-episode psychosis. *British Journal of Clinical Psychology 38*(2), 181-194.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. Psychological Medicine, 42(9), 1879.

Kurtz, M. M., & Richardson, C. L. (2012). Social cognitive training for schizophrenia: a meta-analytic investigation of controlled research. *Schizophrenia Bulletin 38*(5), 1092-1104.

Kurtz, M. M., Gagen, E., Rocha, N. B., Machado, S., & Penn, D. L. (2016). Comprehensive treatments for social cognitive deficits in schizophrenia: A critical review and effect-size analysis of controlled studies. *Clinical Psychology Review 43*, 80-89.

Lahera, G., Herrera, S., Reinares, M., Benito, A., Rullas, M., González-Cases, J., & Vieta, E. (2015). Hostile attributions in bipolar disorder and schizophrenia contribute to poor social functioning. *Acta Psychiatrica Scandinavica 131*(6), 472-482.

Lako, I. M., Bruggeman, R., Knegtering, H., Wiersma, D., Schoevers, R. A., Slooff, C. J., & Taxis, K. (2012). A systematic review of instruments to measure depressive symptoms in patients with schizophrenia. *Journal of Affective Disorders*, *140*(1), 38-47.

Langdon, R., & Coltheart, M. (2000). The cognitive neuropsychology of delusions. *Mind & Language*, *15*(1), 184-218.

Langdon, R., Corner, T., McLaren, J., Ward, P. B., & Coltheart, M. (2006). Externalizing and personalizing biases in persecutory delusions: the relationship with poor insight and theory-of-mind. *Behaviour Research and Therapy 44*(5), 699-713.

Langdon, R., Still, M., Connors, M. H., Ward, P. B., & Catts, S. V. (2013). Attributional biases, paranoia, and depression in early psychosis. *British Journal of Clinical Psychology 52*(4), 408-423.

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

le Gall, E., Iakimova, G., & Askenazy, F. (2013). Attribution bias in early-onset-schizophrenia: relationship to clinical features. *European Psychiatry*, *28*(S1), 1.

Leifker, F. R., Patterson, T. L., Heaton, R. K., & Harvey, P. D. (2011). Validating measures of real-world outcome: the results of the VALERO expert survey and RAND panel. *Schizophrenia Bulletin 37*(2), 334-343.

Leucht, S., Pitschel-Walz, G., Abraham, D., & Kissling, W. (1999). Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophrenia Research 35*(1), 51-68.

Lincoln, T. M., Mehl, S., Exner, C., Lindenmeyer, J., & Rief, W. (2010). Attributional style and persecutory delusions. Evidence for an event independent and state specific external-personal attribution bias for social situations. *Cognitive Therapy and Research 34*(3), 297-302.

Lyon, H. M., Kaney, S., & Bentall, R. P. (1994). The defensive function of persecutory delusions. Evidence from attribution tasks. *The British Journal of Psychiatry*, *164*(5) 637-646.

Lysaker, P. H., Bob, P., Pec, O., Hamm, J., Kukula, M., Vohs, J., Popolo, R., Salvatore, G. & Dimaggio, G. (2013). Synthetic metacognition as a link between brain and behavior in schizophrenia. *Translational Neuroscience*, *4*(3), 368-377.

McIntosh, L. G., & Park, S. (2014). Social trait judgment and affect recognition from static faces and video vignettes in schizophrenia. *Schizophrenia Research*, *158*(1), 170-175.

Majadas, S., Olivares, J., Galan, J., & Diez, T. (2012). Prevalence of depression and its relationship with other clinical characteristics in a sample of patients with stable schizophrenia. *Comprehensive Psychiatry 53*(2), 145-151.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology 33*, 101–121

Mancuso, F., Horan, W. P., Kern, R. S., & Green, M. F. (2011). Social cognition in psychosis: multidimensional structure, clinical correlates, and relationship with functional outcome. *Schizophrenia Research 125*(2), 143-151.

Mansfield, J. L., & Wade, T. (2000). Assessing the relevance of the hopelessness theory of depression to women with disordered eating. *International Journal of Eating Disorders 28*(1), 113-119.

Martin, J. A., & Penn, D. L. (2002). Attributional style in schizophrenia: an investigation in outpatients with and without persecutory delusions. *Schizophrenia Bulletin 28*(1), 131-141.

McDonald, S., Flanagan, S., Martin, I., & Saunders, C. (2004). The ecological validity of TASIT: A test of social perception. *Neuropsychological Rehabilitation*, *14*(3), 285-302.

McGurk, S. R., Twamley, E. W., Sitzer, D. I., McHugo, G. J., & Mueser, K. T. (2007). A meta-analysis of cognitive remediation in schizophrenia. *American Journal of Psychiatry 164*(12), 1791-1802.

Mehl, S., Rief, W., Lüllmann, E., Ziegler, M., Müller, M. J., & Lincoln, T. M. (2010). Implicit attributional style revisited: evidence for a state-specific "self-decreasing" implicit attributional style in patients with persecutory delusions. *Cognitive Neuropsychiatry 15*(5), 451-476.

Mehl, S., Landsberg, M. W., Schmidt, A. C., Cabanis, M., Bechdolf, A., Herrlich, J., Loos-Janowiak, S., Kircher, T., Kiszenow, S., Klingberg, S., Kommescher, M., Moritz, S., Muller, B.W., Sartory, G., Wiedemann, G., Wittorf, A., Wolwer, W. & Wagner, M. (2014). Why do bad things happen to me? Attributional style, depressed mood, and persecutory delusions in patients with schizophrenia. *Schizophrenia Bulletin 40*(6), 1338-1346.

Melo, S., Corcoran, R., Shryane, N., & Bentall, R. P. (2009). The persecution and deservedness scale. Psychology and Psychotherapy: Theory, Research and Practice, 82(3), 247-260.

Montag, C., Neuhaus, K., Lehmann, A., Krüger, K., Dziobek, I., Heekeren, H. R., ... & Gallinat, J. (2012). Subtle deficits of cognitive theory of mind in unaffected first-degree relatives of schizophrenia patients. *European Archives of Psychiatry and Clinical Neuroscience*, *262*(3), 217-226.

Moritz, S., Woodward, T. S., Burlon, M., Braus, D. F., & Andresen, B. (2007). Attributional style in schizophrenia: evidence for a decreased sense of self-causation in currently paranoid patients. *Cognitive Therapy and Research 31*(3), 371-383.

Moritz, S., Burnette, P., Sperber, S., Köther, U., Hagemann-Goebel, M., Hartmann, M., & Lincoln, T. M. (2011). Elucidating the black box from stress to paranoia. *Schizophrenia Bulletin 37*(6), 1311-1317.

Moritz, S., Köther, U., Hartmann, M., & Lincoln, T. M. (2015). Stress is a bad advisor. Stress primes poor decision making in deluded psychotic patients. *European Archives of Psychiatry and Clinical Neuroscience*. *Online Access.*

Moritz, S., Werner, R., & Collani, G. V. (2006). The inferiority complex in paranoia readdressed: a study with the Implicit Association Test. *Cognitive Neuropsychiatry 11*(4), 402-415.

Moritz, S., Veckenstedt, R., Hottenrott, B., Woodward, T. S., Randjbar, S., & Lincoln, T. M. (2010). Different sides of the same coin? Intercorrelations of cognitive biases in schizophrenia. *Cognitive Neuropsychiatry*, *15*(4), 406-421.

Morrison, S. C., & Cohen, A. S. (2014). The moderating effects of perceived intentionality: exploring the relationships between ideas of reference, paranoia and social anxiety in schizotypy. *Cognitive Neuropsychiatry 19*(6), 527-539.

Patterson, T. L., Moscona, S., McKibbin, C. L., Davidson, K., & Jeste, D. V. (2001). Social skills performance assessment among older patients with schizophrenia. *Schizophrenia Research*, *48*(2), 351-360.

Payne, B. K. (2001). Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181-192.

Penn, D. L., Corrigan, P. W., Bentall, R. P., Racenstein, J., & Newman, L. (1997). Social cognition in schizophrenia. *Psychological Bulletin 121*(1), 114-132.

Peters, E. R., Moritz, S., Schwannauer, M., Wiseman, Z., Greenwood, K. E., Scott, J. & Garety, P. A. (2014). Cognitive Biases Questionnaire for Psychosis. *Schizophrenia Bulletin 40*(2), 300-313.

Peterson, C., Semmel, A., Von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. (1982). The attributional style questionnaire. *Cognitive Therapy and Research 6*(3), 287-299.

Peyroux, E., Strickland, B., Tapiero, I., & Franck, N. (2014). The intentionality bias in schizophrenia. *Psychiatry Research 219*(3), 426-430.

Pinkham, A. E., Hopfinger, J. B., Pelphrey, K. A., Piven, J., & Penn, D. L. (2008). Neural bases for impaired social cognition in schizophrenia and autism spectrum disorders. *Schizophrenia Research*, *99*(1), 164-175.

Pinkham, A. E., Harvey, P. D., & Penn, D. L. (2016). Paranoid individuals with schizophrenia show greater social cognitive bias and worse social functioning than non-paranoid individuals with schizophrenia. *Schizophrenia Research: Cognition 3*, 33-38.

Pinkham, A. E., Penn, D. L., Green, M. F., Buck, B., Healey, K., & Harvey, P. D. (2015). The Social Cognition Psychometric Evaluation Study: results of the expert survey and RAND panel. *Schizophrenia Bulletin 40*(4), 813-823.

Pinkham, A. E., Penn, D. L., Green, M. F., & Harvey, P. D. (2015). Social cognition psychometric evaluation: Results of the initial psychometric study. *Schizophrenia Bulletin*, 494-504.

Randall, F., Corcoran, R., Day, J., & Bentall, R. (2003). Attention, theory of mind, and causal attributions in people with persecutory delusions: a preliminary investigation. *Cognitive Neuropsychiatry 8*(4), 287-294.

Randjbar, S., Veckenstedt, R., Vitzthum, F., Hottenrott, B., & Moritz, S. (2011). Attributional biases in paranoid schizophrenia: Further evidence for a decreased sense of self-causation in paranoia. *Psychosis 3*(1), 74-85.

Roberts, D. L., & Penn, D. L. (2009). Social cognition and interaction training (SCIT) for outpatients with schizophrenia: a preliminary study. *Psychiatry Research 166*(2), 141-147.

Roberts, D. L., Penn, D. L., Labate, D., Margolis, S. A., & Sterne, A. (2010). Transportability and feasibility of Social Cognition And Interaction Training (SCIT) in community settings. *Behavioural nd Cognitive Psychotherapy 38*(1), 35-47.

Roberts, D. & Pinkham, A.E. (2013). The future of social cognition in schizophrenia: Implications from the normative literature. In: *Social Cognition in Schizophrenia.* Penn, D.L. & Roberts, D. (Eds.) Oxford: Oxford University Press.

Rocha, N. B., & Queirós, C. (2013). Metacognitive and social cognition training (MSCT) in schizophrenia: A preliminary efficacy study. *Schizophrenia Research*, *150*(1), 64-68.

Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition*, *108*(3), 771-780.

Saha, S., Chant, D., Welham, J., & McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. *PLoS Medicine 2*(5), 413-433.

Sanjuán, P., & Magallares, A. (2009). A longitudinal study of the negative explanatory style and attributions of uncontrollability as predictors of depressive symptoms. *Personality and Individual Differences 46*(7), 714-718.

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. *Schizophrenia Bulletin 39*(5), 979-992.

Schneider, L. C., & Struening, E. L. (1983, September). SLOF: a behavioral rating scale for assessing the mentally ill. In *Social Work Research and Abstracts* (Vol. 19, No. 3, pp. 9-21). Oxford University Press.

Schoenherr, S. J., Brown, R. T., Baldwin, K., & Kaslow, N. J. (1992). Attributional styles and psychopathology in pediatric chronic-illness. *Journal of Clinical Child Psychology 21*(4), 380-387.

Sergi, M. J., Rassovsky, Y., Nuechterlein, K. H., & Green, M. F. (2006). Social perception as a mediator of the influence of early visual processing on functional status in schizophrenia. *American Journal of Psychiatry 163*(3), 448-454.

Sergi, M., Green, M., Widmark, C., Reist, C., Erhart, S., Braff, D., Kee, K.S., Marder, S.R. & Mintz, J. (2007). Social cognition and neurocognition: effects of risperidone, olanzapine, and haloperidol. *American Journal of Psychiatry 164*(10), 1585-1592.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*(1), 131-142.

Swartz, M. S., Perkins, D. O., Stroup, T. S., Davis, S. M., Capuano, G., Rosenheck, R. A., Reimherr, F., McGee, M.F., Keefe, R.S.E., McEvoy, J.P., Hsiao, J.K., & Lieberman, J. A. (2007). Effects of antipsychotic medications on psychosocial functioning in patients with chronic schizophrenia: findings from the NIMH CATIE study. *The American Journal of Psychiatry 164*(3), 428-436.

Sweeney, P. D., Anderson, K., & Bailey, S. (1986). Attributional style in depression: a meta-analytic review. *Journal of Personality and Social Psychology 50*(5), 974-992.

Tas, C., Danaci, A. E., Cubukcuoglu, Z., & Brüne, M. (2012). Impact of family involvement on social cognition training in clinically stable outpatients with schizophrenia—a randomized pilot study. *Psychiatry Research 195*(1), 32-38.

Thewissen, V., Bentall, R. P., Oorschot, M., van Lierop, T., van Os, J., & Myin-Germeys, I. (2011). Emotions, self-esteem, and paranoid episodes: An experience sampling study. *British Journal of Clinical Psychology*, *50*(2), 178-195.

Twamley, E. W., Jeste, D. V., & Bellack, A. S. (2003). A review of cognitive training in schizophrenia. *Schizophrenia Bulletin 29*(2), 359-382.

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. *Schizophrenia Research 103*(1), 257-265.

Vassos, E., Pedersen, C. B., Murray, R. M., Collier, D. A., & Lewis, C. M. (2012). Meta-analysis of the association of urbanicity with schizophrenia. *Schizophrenia Bulletin*, *38*(6), 1118-1123.

Velligan, D. I., Mahurin, R. K., Diamond, P. L., Hazleton, B. C., Eckert, S. L., & Miller, A. L. (1997). The functional significance of symptomatology and cognitive function in schizophrenia. *Schizophrenia Research 25*(1), 21-31.

Waldheter, E. J., Jones, N. T., Johnson, E. R., & Penn, D. L. (2005). Utility of social cognition and insight in the prediction of inpatient violence among individuals with a severe mental illness. *The Journal of Nervous and Mental Disease*, *193*(9), 609-618.

Wang, Y., Roberts, D. L., Xu, B., Cao, R., Yan, M., & Jiang, Q. (2013). Social cognition and interaction training for patients with stable schizophrenia in Chinese community settings. *Psychiatry Research 210*(3), 751-755.

Wassink, T. H., Flaum, M., Nopoulos, P., & Andreasen, N. C. (1999). Prevalence of depressive symptoms early in the course of schizophrenia. *American Journal of Psychiatry 156*(2), 315-316.

Wegner, D. M. (2002). *The illusion of conscious will.* Cambridge, MA: MIT Press.

Wiens, A. N., Bryan, J. E., & Crossen, J. R. (1993). Estimating WAIS-R FSIQ from the national adult reading test-revised in normal subjects. *The Clinical Neuropsychologist 7*(1), 70-84.