HHS Public Access

Author manuscript

J Psychiatr Res. Author manuscript; available in PMC 2022 August 01.

Published in final edited form as:

J Psychiatr Res. 2021 August ; 140: 436–442. doi:10.1016/j.jpsychires.2021.06.016.

Using a Meta-Cognitive Wisconsin Card Sorting Test to Measure Introspective Accuracy and Biases in Schizophrenia and Bipolar Disorder

Bianca A. Tercero[a], **Michelle M. Perez**[a], **Noreen Mohsin**[a], **Raeanne C. Moore**[b,c], **Colin A. Depp**[b,c], **Robert A. Ackerman**[d], **Amy E. Pinkham**[d,e], **Philip D. Harvey**[a,f]

[a]Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, 1120 NW 14th Street, Suite 1450, Miami, FL 33136 USA

[b]Department of Psychiatry, University of California, San Diego, California, USA

[c]VA San Diego Healthcare System, San Diego, California, USA

[d]School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, TX, USA

[e]Department of Psychiatry, University of Texas Southwestern Medical School, Dallas, TX, USA

[f]Research Service, Miami VA Healthcare System, Miami, FL, USA

Abstract

People with schizophrenia (SCZ) and bipolar disorder (BD) have challenges in self-evaluation of their cognitive and functional performance (introspective accuracy). They also manifest response biases, with tendencies toward overestimation. This study aimed to examine objective test performance, momentary judgments of performance, momentary confidence, and subsequent global judgments of performance on a metacognitive version of the Wisconsin Card Sorting Test (WCST). This sample included 99 participants with SCZ and 67 with BD. After each of the 64 WCST trials, participants reported whether they believed their sort was correct and how confident they were in that judgment, they then received performance feedback. After completion of the entire task, participants generated a global performance judgment. On average, the SCZ group got 31 sorts correct, reporting being correct on 49 whereas the BD group got 37 trials correct but reported being correct on 53. For participants with BD, sorting performance correlated with trial x trial accuracy judgments, confidence, and predicted global judgments. For SCZ participants, performance minimally correlated with trial x trial accuracy judgments, confidence, and global judgments, while trial x trial confidence was strongly associated with trial x trial accuracy

*Corresponding Author: Philip D. Harvey, PhD, Miller School of Medicine, University of Miami, 1120 NW 14th Street, Suite 1450, Miami, FL 33136 US (telephone: 305-243-4094; fax: 305-243-1619; pharvey@miami.edu.

Author contributions.

Drs. Harvey, Pinkham, Moore, Ackerman, and Depp designed the study. Dr. Harvey ran the data analyses in consultation with Ms. Mohsin, Perez, and Tercero. Ms. Mohsin, Perez, and Tercero wrote the first draft of the manuscript. All authors have reviewed and edited the paper and approve the final version.

judgments (r=.58). Our findings suggest that confidence in participants with BD is correlated with task performance, whereas in SCZ confidence was entirely associated with self-generated performance judgments. SCZ participants manifested challenges with utilization of feedback. Global judgments of performance were predicted by task performance and confidence for BD participants, with performance and confidence judgments occurring prior to generation of the global performance judgments.

Keywords

Schizophrenia; Bipolar Disorder; Executive Functioning; Introspective Accuracy; Introspective Bias

1. Introduction

Introspective Accuracy (IA) is defined as an individual's ability to evaluate their own abilities, skills, everyday functional performance, and decision-making skills (Harvey and Pinkham, 2015), referenced by objective information. While IA refers to the degree of accuracy in self assessments, it does not index the nature of response biases in cases of inaccuracy. The direction of bias in case of failures in IA can be overestimation or underestimation and is referred to as Introspective Bias (IB). These two concepts are related because if an individual has impaired self-assessment (IA) they could either over or underestimate (IB) their abilities (Silberstein & Harvey, 2019). Harvey and Pinkham (2015) evaluated the applicability of judgements of IA to clinical, functional and social cognitive domains. Although introspective accuracy encompasses various domains, it is not necessarily a global trait of an individual, and an individual may have awareness or biases in one domain but not in others (Gilleen et al., 2011).

Prior studies have reported that individuals with schizophrenia (SCZ) often lack awareness of their clinical symptoms. Interestingly, Medalia et al. (2009) found that participants with schizophrenia had relatively less awareness of their level of neurocognitive deficit compared to awareness of symptoms. In their sample, 70% had full insight into their clinical symptoms while only 27% demonstrated awareness of their neurocognitive deficits and the direction of bias was on overestimation bias. It is important to note that participants with SCZ who lack clinical insight can characterize delusions and hallucinations on the part of others as implausible experiences, such as in the classic study by Rokeach (1964). Recent studies have also shown that impaired social cognition and functioning is not related to lack of understanding of socially normative behavior (Langdon et al., 2014).

Several previous studies have demonstrated that individuals with SCZ manifest a strong overestimation bias during performance-based assessments. Pinkham et al. (2018) reported that higher levels of confidence in social cognitive decision making across two different tests accounted for 14% and 17% of the variance in poorer social outcomes, while accuracy on those same tasks accounted for less than 1% of the variance in social functioning. Jones et al. (2019) found that participants with SCZ overestimated their social cognitive performance at every level of accuracy. Of note, 18% of participants with SCZ reported with 100% confidence that they were correct on every item of the task, while these participants were the

poorest performers. Badal et al. (2021) used machine learning strategies to compare social cognitive performance, confidence, and effort in order to identify the optimally discriminating elements of task performance to differentiate healthy people and participants with schizophrenia. Confidence ratings were twice as efficient as task performance for discriminating the groups Perez et al. (2020) found that overconfidence in social cognitive performance predicted poorer performance on multiple tasks of neurocognition and social cognition. Overconfidence may lead to worse performance by inducing failures to adjust efforts when faced with tasks varying in difficulty (Cornacchio et al., 2017). Another possibility for the relationship between overconfidence and poor performance in participants with SCZ is a bias toward using self-generated rather than objectively obtained information, which we explore further in this study.

Comprehensive models of self-assessment in neuropsychiatric conditions have addressed several elements of metacognition. Essentially synonymous to IA, one of these processes is referred to as “metacognitive monitoring accuracy” (Koren et al., 2005). Other elements of these more comprehensive models include 'monitoring resolution' (the correlation within each subject between the confidence in each given sort and the actual correctness of this sort), and 'control sensitivity' (the correlation within each subject between the confidence in each given sort and the judgment of whether it is correct or not). However, these constructs are stretched in situations where participants report that they believe that are correct on all task items and have 100% confidence in the accuracy of judgments. These models make a critical assumption, that participants use all sources of information and attempt to generate an objectively derived estimate of their performance.

These multiple impairments in self-awareness and positive response biases have functional implications. Individuals with SCZ may experience interpersonal challenges because overconfidence leads to a disconnect between how they perceive themselves and how others view them (Lysaker et al., 1998). For example, Gould et al. (2015) found that overestimation of cognitive and functional capacity performance was more strongly correlated with everyday function than with performance on the ability measures. Silberstein et al. (2018) similarly found that overly positive introspective biases when judging social cognitive abilities was a better predictor of social everyday disability than actual performance on social cognitive tasks.

Impairments in self-assessment are found in other conditions in addition to SCZ (Varga et al., 2007). There has been considerably less research on all forms of IA in BD compared to schizophrenia. Pini et al. (2001) reported that people with SCZ did not differ from people with BD in terms of clinical insight. Strassnig et al. (2018b) found that in participants with BD, those who were unemployed or not living independently did not self-report more disability than those with current milestone achievements in these two domains.

In our earlier studies (e.g., Cornacchio et al., 2017; Gould et al., 2015; Jones et al., 2019) participants did not receive trial x trial feedback on their performance, although they rated their confidence in the accuracy of their proximal response on each trial. Thus, in those studies, there was no possibility of self-correcting confidence or later accuracy judgments following feedback regarding performance. The present study adopted a different strategy.

We adapted a multi-trial executive functioning test, the metacognitive Wisconsin Card Sorting Test (WCST; Koren et al., 2004), that we had previously employed (Gould et al., 2015). We made some modifications to the test, in that after each of the 64 sorts, the participant was asked to make an accuracy judgment about their sort. They were then asked about their confidence in this accuracy judgment and were then provided performance feedback (correct/incorrect). At the end of the assessment, they were asked to provide a global judgment about their performance on the task and then asked to provide a similar global judgment about how a "typical person" would do on this task.

In these analyses we intentionally did not relate item x item accuracy to item x item judgments and item x item confidence, seeking instead to see if global indices of performance, judgment, and confidence predicted global judgments regarding cognitive ability. We also assessed psychosis, negative symptoms of reduced emotional experience, and depression. Several models of the origin of psychotic symptoms implicate impaired momentary judgements as a possible origin and very low levels of depression have been reported to correlate with a highly positive functional IB in participants with SCZ (Harvey et al., 2017b; Harvey et al., 2019; Siu, et al., 2015). In contrast, in a pilot study of participants with BD, higher levels of depression were associated with negative IB compared to clinician ratings (Harvey et al., 2015).

We had several hypotheses. We hypothesized that participants in both groups would manifest impaired introspective accuracy combined with a positive introspective bias, leading to overestimations of total accuracy. We hypothesized that higher average confidence ratings would be related to poorer overall performance, in line with our previous findings in SCZ participants. We hypothesized that that average confidence ratings across trials would be correlated more strongly with total scores for accuracy judgments than with actual performance. We also hypothesized that global judgments about task performance would be correlated with total scores for confidence and total scores on accuracy judgments, but not by total performance. In terms of symptom correlations, we expected that very low levels of clinically rated depression would correlate with positive IB and that psychotic symptoms would correlate with impairments in IA.

2. Methods

2.1 Participants

Participants who met DSM-V criteria for Schizophrenia, Schizoaffective Disorder, or Bipolar Disorder (I or II), with or without current or previous psychotic symptoms, participated in this study. All schizophrenia spectrum patients were grouped into a single group, as were the BD participants. They were recruited at three different sites: The University of Miami Miller School of Medicine (UM), the University of California San Diego (UCSD), and The University of Texas at Dallas (UTD). UM participants were recruited from the Jackson Memorial Hospital-University of Miami Medical Center and the Miami VA Medical Center. UCSD participants were recruited from the UCSD Outpatient Psychiatric Services clinic, a large public mental health clinic, the San Diego VA Medical Center, and other local community clinics and by word of mouth. UTD participants were recruited primarily from Metrocare Services, a non-profit mental health services

organization in Dallas County, TX, and from other local clinics. The study was approved by each University's respective Institutional Review Board, and all participants provided written informed consent. Diagnostic information was collected by trained interviewers using the Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) and the psychosis module of the Structured Clinical Interview for DSM Disorders-5 (SCID-5; First et al., 2015), and a local consensus procedure was used to generate final diagnoses.

2.2 Inclusion/Exclusion Criteria

To be eligible, participants had to meet criteria for one of the disorders mentioned above. Individuals with BD also had to meet a staging model severity of 3 or higher, indicating at least one mood episode recurrence or incomplete remission from a first-episode (Frank et al., 2014). Participants were also required to be clinically stable (i.e. no hospitalizations) for a minimum of 6 weeks and to be on a stable medication regimen for a minimum of 6 weeks with no dose changes >20% for a minimum of 2 weeks. All antipsychotics or antipsychotic combinations were accepted.

For participants in both diagnostic groups exclusion criteria included: (1) history of or current medical or neurological disorders that may affect brain functioning (e.g., CNS tumors, seizures, or loss of consciousness for over 15 minutes), (2) history of or current intellectual disability (IQ<70) or pervasive developmental disorder according to the DSM-5 criteria, (3) presence of substance use disorder not in remission for at least six months, (4) visual or hearing impairments that interfere with assessment, and (5) lack of proficiency in English. Participants with a Wide Range Achievement Test-3rd edition (WRAT-3; Jastak, 1993) grade equivalent score of less than 8th grade were also not enrolled.

2.3 Clinical Assessments

2.3.1 Depression Symptoms.—

We used a common clinician rated depression assessment, the Montgomery-Asberg Depression rating scale (MADRS; Montgomery and Asberg, 1979). MADRS ratings were generated on the day of the cognitive assessment procedure.

2.3.2 Psychotic Symptoms.—

Severity of symptoms was evaluated with the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987), which was administered in its entirety by trained raters and on the same day as the MADRS and the testing procedure. The PANSS consists of 30 items with 3 subscales: positive symptoms, negative symptoms, and general psychopathology. Each item was scored on a 7-point (1–7) Likert scale. These raters had extensive experience in other studies of participants with severe mental illness and were trained to high reliability for both the MADRS and the PANSS (ICC>.80) by the study PI (Pinkham). For the MADRS we used the total score, and the primary PANSS symptoms of interest were the psychosis items generated by the Marder et al. (1997) factor analysis: P1, P2, P3, P5, P6, G9.

2.3.3 Negative Symptoms.—

Khan and colleagues (2017) generated a two-factor model of negative symptoms measured by the PANSS, identifying dimensions of expressive deficits and experiential deficits. This model is clinically relevant as the reduced emotional

experience factor has been shown to predict variance in everyday functioning in several different samples (Harvey et al., 2017a; Strassnig et al., 2018a) and to respond to pharmacological treatment (Harvey et al., 2020). The items in the *PANSS Reduced Emotional Experience* factor are: Emotional Withdrawal (N2), Passive/Apathetic Social Withdrawal (N4) and Active Social Avoidance (G16).

2.3.4 Metacognitive Wisconsin Card Sorting Test.—

This test was similar to that developed by Koren et al. (2004). There were 64 sorts to be performed. After each sort, to measure IA, the participant was asked "Did you get it correct?" and answered with a yes/no response. Then the participants were asked to provide a confidence judgment on a 1–5 (1=0%, 5=100%) scale as to their confidence in the correctness of their accuracy judgment. Participants were provided feedback about their response on a yes/no basis after they had provided both accuracy and confidence judgments. After completion of this assessment, which was administered as part of an array of neurocognitive, social cognitive, social competence, and functional capacity measures to be reported later, they were asked "How did you do on this test?", rated with a 1–100 rating, indexed from 1 (very poorly), 50 (about like the typical person), 100 (perfect). A final question was "How would the typical person do on this test?", also rated with a 1–100 scale. Thus, there were 5 dependent variables, three of which were collected on a trial x trial basis: correct sorts (0–64), accuracy judgments regarding correct sorts (0–64), mean trial-by-trial confidence ratings (1–5 range) and two variables collected afterwards: global assessment of personal performance referenced to the typical person (1–100 range), and estimates of the "typical person's" performance (1–100 range). We also calculated a difference score reflecting IA and IB, by subtracting the number of trials on which the participant reported that they were correct from their actual performance.

2.4 Data Analyses.

Statistical Analyses were performed using IBM SPSS version 26. We used t-tests to compare the three clinical symptom variables and scores on the 5 IA variables as well as the IB difference score between participants with SCZ and BD. We used a single sample t-test within each sample to examine the difference between WCST Performance and Trial x trial accuracy judgments to see if these differences were significantly different from zero and in which direction.

We then computed Pearson product-moment correlations, in the two groups separately, between the 5 WCST variables, the two difference scores (Performance judgments, Global judgments), psychotic symptoms, depression, and reduced emotional experience. Using Fisher's r to Z transformation, we compared the size of the correlations between WCST performance and trial x trial accuracy judgments and trial x trial confidence across the two groups. Finally, we constructed a hierarchical regression model, wherein we predicted average confidence scores with WCST performance and trial-by-trial self-reported performance. This model was designed to examine correlates of confidence ratings. Then we predicted global performance judgments with WCST performance, trial-by-trial accuracy judgments, and trial-by-trial confidence ratings.

3. Results.

Descriptive information on the individuals with BD (n=67) and SCZ (n=99) is presented in Table 1. Participants with BD had more education and higher WRAT-3 reading scores than the participants with SCZ. In terms of functional milestones, participants with BD were more likely to have ever been married or equivalent, with no significant differences in current residential or employment status. Participants with SCZ had higher scores on psychosis and reduced emotional experience, and there was no difference in MADRS scores between the groups.

The 5 WCST performance and self-assessment variables are presented in Table 2. As can be seen, participants with BD out-performed the participants with SCZ on the number of correct sorts. However, were no group differences in trial x trial accuracy judgments, trial-by-trial confidence, global judgments of personal performance, and global judgments of difficulty or the two difference scores. For both the SCZ and BD participants, the number of positive accuracy judgments was significantly higher than WCST performance, both $t>9.57$, both $p<.001$, reflecting statistically significant impairment in introspective accuracy and a statistically significant positive introspective bias.

As there were sex and race differences between the samples, we examined 5 all of the main outcomes variables with sex x Diagnosis and race x diagnosis analyses of variance. All effects of sex and sex interactions were non-significant, all $F<3.25$, all $p>.07$. All effects of racial status and racial status were similarly non-significant, all $F<1.95$, all $p>.09$

3.1 Pearson Correlations

In contrast to our hypotheses, none of the correlations between the WCST Variables, MADRS scores, Psychosis, and Reduced Emotional experience were statistically significant in either group, all $r<.18$, all $p>.08$. Intercorrelations for the WCST variables are presented in Table 3. For participants with SCZ, the number of correct WCST sorts was significantly, but modestly, correlated with trial-by-trial accuracy judgments, impressions of personal global performance, and impressions of general difficulty, but was unrelated to trial-by-trial confidence. In contrast, trial-by-trial accuracy judgments were more substantially correlated with trial x trial confidence and global judgments. For participants with BD, correct WCST sorts were correlated with all other variables, including confidence. Accuracy judgments were correlated with all of the other variables other than judgments of task difficulty for the general population. Trial-by-trial confidence was correlated with all other variables as was global personal judgments regarding performance.

Overestimation of performance was correlated with poorer accuracy in the SCZ participants and was positively correlated with trial x trial accuracy judgments, trial x trial confidence, and global judgments of performance. In the BD participants the correlation between overestimation and reduced accuracy was significant but smaller. However, trial x trial confidence was not correlated with over-estimation global performance judgments were negatively related to overestimated performance on a trial x trial basis

Using Fisher's r to Z transformation, we compared the size of the correlations between WCST performance and trial x trial accuracy judgments and trial x trial confidence across the two groups. The correlation between WCST performance and a trial x trial accuracy judgment were significantly higher in the BD participants than in the SCZ participants, z=2.01, p<.05, and the correlation between WCST performance and trial x trial confidence was also significantly higher in the BD participants, z=2.69, p=.004. The difference in the size of the correlations between greater overestimation and lower accuracy was significant across the two samples, z=1.99, p<.05, and difference between the correlations between overestimation and global judgments of performance was significant as well, z=6.18, p<.001.

3.2 Regression Models

The regression models predicting trial-by-trial confidence and global judgments of personal performance are presented in Table 4. As can be seen in the table, in participants with SCZ the regression model examining trial-by-trial confidence ratings was statistically significant, F (1,97)=48.49, p<.001. Accuracy judgments entered the model, accounting for 33% of the variance (p<.001), but WCST performance did not enter the model. For participants with BD, the model was also significant overall, F (2,64) =13.00, p<.001. WCST performance entered the model first, accounting for 24% of the variance (p<.001), and trial x trial accuracy judgments entered second, accounting for 5% of the variance (p=.038).

For prediction of global performance judgments in the participants with SCZ, the overall model was significant, F (2,94) =38.95, p<.001. Trial x trial confidence entered the model first, accounting for 26% of the variance (p<.001), while trial x trial accuracy judgments entered second, accounting for 20% of the variance (p<.001). For the participants with BD, the regression was also significant, F (2,64) =36.96, p<.001. All three predictors entered the model, with WCST performance accounting for 57% of the variance (p<.001), trial x trial confidence ratings accounting for an additional 4% (p=.012), and trial x trial accuracy judgments accounting for an additional 2% (p=.045).

4. Discussion

Overall, participants with SCZ and BD tended to overestimate their cognitive test performance by about 50% on a momentary basis. This difference between task performance and trial x trial judgments was statistically significant in both groups. Furthermore, the level of trial x trial confidence was very high (4 out of 5), given the number of errors in both groups (52% and 42% respectively). The BD participants had significantly better performance on the test than the participants with SCZ, but there were no group differences in the trial-by-trial IA or IB ratings, trial-by-trial confidence, global judgments of personal performance, and global judgments of task difficulty. Thus, while groups differed in performance, they did not differ in IA or IB for momentary assessment of their WCST performance.

The major difference between the groups was in the association of accuracy on the task with correlates of both trial x trial confidence and global judgments regarding their performance. Both groups of participants generated global performance judgments that were correlated on

a zero-order basis with the accuracy of their performance as well as their trial x trial confidence and their trial x trial accuracy judgments. However, the shared variance for performance and accuracy judgments was only 4% for the participants with SCZ, compared to over 25% for the participants with BD. In participants with BD, regression analyses confirmed that WCST performance was the single best predictor of their global judgments, while in the participants with SCZ, the regression analysis suggested that task performance did not contribute to global estimates of ability when trial x trial confidence and trial x trial accuracy judgments were considered. Additionally, trial x trial confidence ratings were not independently related to task performance in the SCZ participants, while for participants with BD, performance on the WCST was the largest predictor of trial-by-trial confidence.

Solving the WCST requires the participant to understand a sorting concept, based on the abstract information provided by each card, and utilize feedback from the outcome of the previous trial. Previous studies on the WCST have suggested that that low scores on this task in participants with severe mental illness typically result from a failure to shift from incorrect category choices after receiving error feedback (Goldberg and Weinberger, 1988). Furthermore, verbal working memory performance is associated with WCST performance, suggesting that failures to use feedback may be due to failure to remember the feedback (Gold et al., 1997). Participants with BD apparently were incorporating the feedback provided, since their WCST performance was significantly better than the participants with SCZ and was also significantly more associated with trial x trial confidence and accuracy judgements. For patients with BD, this incorporation of feedback is reflected by the fact that the **more** that they overestimated their performance on a trial x trial basis, the **poorer** were their global judgments of their performance. In contrast, the patients with SCZ did not appear to incorporate the provided feedback into their momentary judgements, their trial x trial confidence, or their global judgments of their performance. Previously, Prentice et al. (2008) found that in patients with SCZ, impaired use of feedback is evident as early as the first four trials of the WCST. In the Prentice et al. study, participants with SCZ were less accurate than healthy individuals, and performance in those early sorts predicted their overall task performance

In our study, the participants with schizophrenia apparently remembered their impressions regarding their trial x trial performance, because their judgments about their global performance, generated after the task was completed, were substantially correlated with these indices. This is consistent with previous studies where participants with SCZ seem biased toward recollection of self-generated responses compared to those that originated from external sources (Vinogradov et al., 1997). Related studies have found that participants with SCZ arrive at judgements with less evidence and are more likely to be convinced of the accuracy of their judgements compared to healthy participants. Consequently, misjudgments may be exaggerated when conclusions are reached rapidly and firmly held, as reported by Moritz et al. (2006, 2012, 2014, 2015) and this effect may be increased when self-generated versus other generated information needs to be discriminated.

Participants with BD over estimated their performance on a trial x trial basis over the duration of the assessment period. Although they appeared to incorporate feedback from task performance into their subsequent global judgments of their performance, they still

faced challenges in the momentary ability to render accurate performance judgments. Thus, tendencies toward responding affirmatively when incorrect may reflect the impulsive cognitive style commonly reported in patients with BD.

Participants with SCZ have been previously reported to fail to adjust their effort to match task difficulty. Cornacchio et al. (2017) proposed that the source of this failure to adjust is that participants with SCZ may have a universal challenge in estimating the difficulty of tasks. In this study, task difficulty did not vary over time, but people with SCZ appeared to be less able to utilize performance feedback to improve their performance. The inability to use feedback among individuals with SCZ may also identify a mechanism of impaired IA in that tendencies to overestimate abilities will not be corrected in the face of contradictory evidence.

A further important finding from this study is the proportion of participants with schizophrenia who generated trial x trial performance estimates that were extraordinarily high, with accompanying confidence estimates. Fifty percent of the participants with SCZ reported that they were correct 75% of the time or more and 12 participants (13%) reported that they were correct on 100% of their sorts. Confidence ratings for the 50% of participants with SCZ who reported that were correct on 75% or more of their WCST sorts averaged 4.7 on a 5-point scale; not one of these participants had a mean confidence rating of less than 4.0 (80% or more confident). For the participants with schizophrenia who reported that they were correct on all sorts, their mean performance was consistent with the SCZ group as a whole (M=30.5) and 11/12 had a mean confidence rating of 5/5.

The limitations to this analysis include the participant groups not being of equal sizes with more participants with SCZ than there are participants with BD. There were differences in racial and sex distributions that were not associated with performance on the task variables. Moreover, participants could have a diagnosis of SCZ, schizoaffective disorder, BD (type 1 and type 2) with or without psychotic features. This task was performed in a battery with other tasks which will be reported upon later. Without healthy controls we cannot determine the level of objective impairment in WCST performance, but our primary focus was on self-assessment rather than performance.

5. Conclusions

Previous studies have found that participants with SCZ arrive at judgements with less evidence and are more likely to be convinced of the accuracy of their judgements than healthy participants. Consequently, misjudgments may be exaggerated when conclusions are reached rapidly and firmly held, as reported by Moritz et al. (2006, 2012, 2015). We found that both participants with SCZ and BD show poor IA. In SCZ, global judgements of competence appear to be based on momentary accuracy judgements rather than task performance, despite trial x trial feedback. An over-riding positive introspective bias suggests that feedback regarding performance may not be considered in generation of accuracy judgments. In contrast, participants with BD appeared to generate global judgements regarding their performance that were related to their performance more than their momentary judgments, while adjusting their global judgments based on performance-

related feedback. Thus, interventions to improve IA in SCZ should probably focus on sensitivity to feedback and consideration of external information. Classic studies attempting to remediate WCST deficits in people with schizophrenia have focused on this strategy, but did not address the challenge of failures to use momentary information about performance to adjust response biases while performing the test.

Acknowledgments

This research was supported by NIMH grant RO1MH112620 to Dr. Pinkham.

The data in this study are being deposited in the NIMH RDOC repository. 6 Months after data lock they will be available for public access. In the interim, the authors are happy to share the data that underlie this paper.

Conflict of interest statement

Dr. Raeanne C. Moore is a co-founder of KeyWise AI, Inc. and a consultant for NeuroUX. Dr. Harvey has received consulting fees or travel reimbursements from Alkermes, Bio Excel, Boehringer Ingelheim, Intra-Cellular Therapies, Minerva Pharma, Regeneron, Roche Pharma, and Sunovion Pharma. He receives royalties from the Brief Assessment of Cognition in Schizophrenia and the MATRICS Consensus Battery. He is chief Scientific officer of iFunction, Inc.He had research grants from Takeda and from the Stanley Medical Research Foundation. Dr. Pinkham has served as a consultant for Roche Pharma. The other authors have no potential Biomedical Conflicts of Interest.

References

Badal VD, Depp CA, Hitchcock PF, Penn DL, Harvey PD, & Pinkham AE (2021). Computational methods for integrative evaluation of confidence, accuracy, and reaction time in facial affect recognition in schizophrenia. Schizophr. Res. Cogn, 25, 100196. 10.1016/j.scog.2021.100196 [PubMed: 33996517]

Cornacchio D, Pinkham AE, Penn DL, & Harvey PD, 2017. Self-assessment of social cognitive ability in individuals with schizophrenia: Appraising task difficulty and allocation of effort. Schizophr Res. 179, 85–90. 10.1016/j.schres.2016.09.033. [PubMed: 27693281]

First MB, Williams JBW, Karg RS, Spitzer RL 2015. Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV). Arlington, VA, American Psychiatric Association.

Frank E, Nimgaonkar VL, Phillips ML, & Kupfer DJ, 2015.. All the world's a (clinical) stage: rethinking bipolar disorder from a longitudinal perspective. Molecular psychiatry, 20(1), 23–31. 10.1038/mp.2014.71 [PubMed: 25048003]

Gilleen J, Greenwood K, David AS, 2011. Domains of awareness in schizophrenia. Schizophr Bull. 37(1), 61–72. 10.1093/schbul/sbq100. [PubMed: 20851850]

Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR, 1997. Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. Arch Gen Psychiatry. 54(2), 159–165. 10.1001/archpsyc.1997.01830140071013. [PubMed: 9040284]

Goldberg TE, Weinberger DR, 1988. Probing prefrontal function in schizophrenia with neuropsychological paradigms. Schizophr Bull. 14(2), 179–183. 10.1093/schbul/14.2.179. [PubMed: 3059467]

Gould F, McGuire LS, Durand D, Sabbag S, Larrauri C, Patterson TL, Twamley EW, Harvey PD, 2015. Self-assessment in schizophrenia: Accuracy of evaluation of cognition and everyday functioning. Neuropsychology, 29(5), 675–682. 10.1037/neu0000175. [PubMed: 25643212]

Harvey PD, Deckler E, Jones MT, Jarskog LF, Penn DL, & Pinkham AE, 2019. Autism symptoms, depression, and active social avoidance in schizophrenia: Association with self-reports and informant assessments of everyday functioning. Journal of psychiatric research, 115, 36–42. 10.1016/j.jpsychires.2019.05.010. [PubMed: 31102902]

Harvey PD, Khan A, Keefe R, 2017a. Using the Positive and Negative Syndrome Scale (PANSS) to Define Different Domains of Negative Symptoms: Prediction of Everyday Functioning by

Impairments in Emotional Expression and Emotional Experience. Innov Clin Neurosci, 14(1112), 18–22. [PubMed: 29410933]

Harvey PD, Paschall G, Depp C, 2015. Factors influencing self-assessment of cognition and functioning in bipolar disorder: a preliminary study. Cogn Neuropsychiatry. 20(4), 361–371. doi:10.1080/13546805.2015.1044510. [PubMed: 26057868]

Harvey PD, Pinkham AE, 2015. Impaired self-assessment in schizophrenia: Why patients misjudge their cognition and functioning. Current Psychiatry. 14(4), 53–59.

Harvey PD, Saoud JB, Luthringer R, Moroz S, Blazhevych Y, Stefanescu C, & Davidson M, 2020. Effects of Roluperidone (MIN-101) on two dimensions of the negative symptoms factor score: Reduced emotional experience and reduced emotional expression. Schizophr Res; 215:352–356. 10.1016/j.schres.2019.08.029. [PubMed: 31488314]

Harvey PD, Twamley EW, Pinkham AE, Depp CA, & Patterson TL, 2017b. Depression in Schizophrenia: Associations with Cognition, Functional Capacity, Everyday Functioning, and Self-Assessment. Schizophrenia bulletin, 43(3), 575–582. 10.1093/schbul/sbw103. [PubMed: 27440672]

Corporation IBM, 2020. Statistical package for the social sciences (SPSS) Edition 26, Armonk, NY.

Jastak S, 1993. Wide-Range Achievement Test, 3rd ed. San Antonio, TX, Wide Range, Inc.

Jones MT, Deckler E, Laurrari C, Jarskog LF, Penn DL, Pinkham AE, & Harvey PD, 2019. Confidence, performance, and accuracy of self-assessment of social cognition: A comparison of schizophrenia patients and healthy controls. Schizophrenia research. Cognition, 19, 002–2. 10.1016/j.scog.2019.01.002.

Kay SR, Fiszbein A, & Opler LA, 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia bulletin, 13(2), 261–276. 10.1093/schbul/13.2.261. [PubMed: 3616518]

Khan A, Liharska L, Harvey PD, Atkins A, Ulshen D, & Keefe R, 2017. Negative Symptom Dimensions of the Positive and Negative Syndrome Scale Across Geographical Regions: Implications for Social, Linguistic, and Cultural Consistency. Innovations in clinical neuroscience, 14(11–12), 30–40.

Koren D, Seidman LJ, Poyurovsky M, Goldsmith M, Viksman P, Zichel S, & Klein E, 2004. The neuropsychological basis of insight in first-episode schizophrenia: a pilot metacognitive study. Schizophrenia research, 70(2–3), 195–202. 10.1016/j.schres.2004.02.004. [PubMed: 15329296]

Langdon R, Connors MH, & Connaughton E, 2014. Social cognition and social judgment in schizophrenia. Schizophrenia research. Cognition, 1(4), 171–174. 10.1016/j.scog.2014.10.001.

Lysaker PH, Bell MD, Bryson GJ, & Kaplan E, 1998. Insight and interpersonal function in schizophrenia. The Journal of nervous and mental disease, 186(7), 432–436. 10.1097/00005053-199807000-00008. [PubMed: 9680045]

Marder SR, Davis. JM, Chouinard G, 1997. The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: combined results of the North American trials. J Clin Psychiatry. 58(12), 538–46. 10.4088/JCP.v58n1205. [PubMed: 9448657]

Medalia A, Thysen J, 2009. A comparison of insight into clinical symptoms versus insight into neuro-cognitive symptoms in schizophrenia. Schizophr Res. 118, 134–9. 10.1016/j.schres.2009.09.027. [PubMed: 19840898]

Montgomery SA, Asberg M, 1979. A new depression scale designed to be sensitive to change. Br J Psychiatry; 134:382–389. 10.1192/bjp.134.4.382. [PubMed: 444788]

Moritz S, Woodward TS, Rodriguez-Raecke R, 2006. Participants with schizophrenia do not produce more false memories than controls but are more confident in them. Psychol. Med, 36(5), 659–667. 10.1017/S0033291706007252. [PubMed: 16512973]

Moritz S, Woznica A, Andreou C, Köther U, 2012. Response confidence for emotion perception in schizophrenia using a Continuous Facial Sequence Task. Psychiatr. Res, 200(2–3), 202–207. 10.1016/j.psychres.2012.07.007.

Moritz S, Ramdani N, Klass H, Andreou C, Jungclaussen D, Eifler S, … & Zink M, 2014. Overconfidence in incorrect perceptual judgments in participants with schizophrenia. Schizophrenia Research: Cognition, 1(4), 165–170. 10.1016/j.scog.2014.09.003. [PubMed: 29379749]

Moritz S, Göritz AS, Gallinat J, Schafschetzy M, Van Quaquebeke N, Peters MJ, Creou C, 2015. Subjective competence breeds overconfidence in errors in psychosis. A hubris account of paranoia. J. Behav. Ther. Exper. Psychiat, 48, 118–124. 10.1016/j.jbtep.2015.02.011.

Perez MM, Tercero BA, Penn DL, Pinkham AE, Harvey PD, 2020. Overconfidence in social cognitive decision making: Correlations with social cognitive and neurocognitive performance in participants with schizophrenia and healthy individuals. Schizophr Res. 224, 51–57. 10.1016/j.schres.2020.10.005. [PubMed: 33097367]

Pini S, Cassano GB, Dell'Osso L, & Amador XF, 2001. Insight into illness in schizophrenia, schizoaffective disorder, and mood disorders with psychotic features. American journal of psychiatry. 158(1), 122–125. 10.1176/appi.ajp.158.1.122.

Prentice KJ, Gold JM, & Buchanan RW, 2008. The Wisconsin Card Sorting impairment in schizophrenia is evident in the first four trials. Schizophrenia research. 106(1), 81–87. 10.1016/j.schres.2007.07.015. [PubMed: 17933496]

Rokeach M., 1964. Three Christs of Ypsilanti. New York: Knopf.

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, & Dunbar GC, 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of clinical psychiatry. 59(20), 22–57.

Silberstein J, & Harvey PD, 2019. Impaired introspective accuracy in schizophrenia: an independent predictor of functional outcomes. Cognitive neuropsychiatry, 24(1), 28–39. 10.1080/13546805.2018.1549985. [PubMed: 30477401]

Silberstein JM, Pinkham AE, Penn DL, & Harvey PD, 2018. Self-assessment of social cognitive ability in schizophrenia: Association with social cognitive test performance, informant assessments of social cognitive ability, and everyday outcomes. Schizophrenia research, 199, 75–82. 10.1016/j.schres.2018.04.015. [PubMed: 29673732]

Siu CO, Harvey PD, Agid O, Waye M, Brambilla C, Choi WK, & Remington G, 2015. Insight and subjective measures of quality of life in chronic schizophrenia. Schizophrenia research. Cognition, 2(3), 127–132. 10.1016/j.scog.2015.05.002.

Smith TE, Hull JW, & Santos L,, 1998. The relationship between symptoms and insight in schizophrenia: a longitudinal perspective. Schizophrenia research, 33(1–2), 63–67. 10.1016/s0920-9964(98)00063-2. [PubMed: 9783345]

Strassnig M, Bowie C, Pinkham AE, Penn D, Twamley EW, Patterson TL, & Harvey PD, 2018a. Which levels of cognitive impairments and negative symptoms are related to functional deficits in schizophrenia? Journal of psychiatric research, 104, 124–129. 10.1016/j.jpsychires.2018.06.018. [PubMed: 30029051]

Strassnig M, Kotov R, Fochtmann L, Kalin M, Bromet EJ, & Harvey PD, 2018b. Associations of independent living and labor force participation with impairment indicators in schizophrenia and bipolar disorder at 20-year follow-up. Schizophrenia research, 197,150–155. 10.1016/j.schres.2018.02.009. [PubMed: 29472164]

Varga M, Magnusson A, Flekkøy K, David AS, & Opjordsmoen S, 2007. Clinical and neuropsychological correlates of insight in schizophrenia and bipolar I disorder: does diagnosis matter?. Comprehensive psychiatry, 48(6), 583–591. 10.1016/j.comppsych.2007.06.003. [PubMed: 17954145]

Vinogradov S, Willis-Shore J, Poole JH, Marten E, Ober BA, & Shenaut GK, 1997. Clinical and neurocognitive aspects of source monitoring errors in schizophrenia. The American journal of psychiatry, 154(11), 1530–1537. 10.1176/ajp.154.11.1530. [PubMed: 9356560]

Table 1

Descriptive and Demographic Information on Participants

	Schizophrenia n=99		Bipolar Disorder n=67			
	M	SD	M	SD	t	p
Age	41.98	10.44	39.22	11.75	1.63	.11
Years of Education	12.53	2.32	14.22	2.64	4.42	<.001
Mothers Education	13.05	3.54	13.67	3.67	1.81	.069
WRAT-3- Standard Score	95.42	11.85	102.13	11.70	3.67	<.001
MADRS Score	10.38	10.93	13.26	11.06	2.87	.004
Reduced Emotional Experience	6.48	3.10	4.90	2.28	3.65	<.001
	SCZ		BD	X^2	p	
Sex (% Female)	48		69	8.22	.004	
Racial Status (%)						
Caucasian	32		53	15.27	.009	
African American	54		25			
Asian	2		3			
Native American, Hawaiian, Alaskan	1		1			
Other, Multiple, Unknown	11		12			
Ethnic Status						
Hispanic	24		29	0.64	.42	
Non-Hispanic	76		81			
Ever Married or Equivalent	49		70	7.14	.007	
Financially Responsible	71		70	0.02	.88	
Unemployed for More for more than one year	60		45	2.74	.10	

Table 2

Performance on WCST Variables in Participants with Schizophrenia and Bipolar Illness

	Schizophrenia		Bipolar Illness				
	M	SD	M	SD	t	p	d
	N=99		N=67				
Correct Sorts (out of 64)	30.67	11.61	36.95	11.77	3.39	<.001	.72
Positive Accuracy Judgments (out of 64)	49.44	17.27	53.30	15.56	1.47	.14	.22
Difference of Self-Reported And Correct Sorts	18.77	18.56	16.34	13.98	0.91	.37	.14
Mean Confidence Rating: Trial × trial (1–5 range)	3.97	0.77	4.01	0.74	0.39	.70	.05
Global Judgment: Personal Performance (0–100)	52.58	24.77	57.95	24.19	1.35	.18	.23
Global Judgment: Performance of General Population	59.39	24.79	59.65	21.21	0.71	.94	.01

Table 3

Pearson Correlations between WCST Variables in the Overall Sample

Participants with Schizophrenia					
	Accuracy Judgments	Confidence	Personal Performance	General Difficulty	Difference of Performance and Accuracy
Correct Sorts	.22 *&	.10 &	.37 ***&	.34 ***	−.42 ***&
Accuracy Judgments	- -	.58 ***	.48 ***	.23 *	.79 ***
Mean Confidence Rating		- -	.51 ***	.28 **	.48 ***
Personal Performance			- -	.54 ***	.60 ***
Participants with Bipolar Illness					
	Accuracy Judgments	Confidence	Personal Performance	General Difficulty	Difference of Performance And Accuracy
Correct Sorts	.50 ***&	.49 ***&	.76 ***&	.29 *	−.28 **&
Accuracy Judgments	- -	.44 ***	.30 *	.21	.69 ***
Mean Confidence Rating		- -	.56 ***	.33 **	.08
Personal Performance			- -	.44 ***	−.30 **

Note.

* p<.05

** p<.01

*** p<.001

& significantly different between groups using Fisher's r to z transformation.

Table 4

Results of Regression analyses Predicting Trial-by-Trial Confidence and Global Performance Judgments

Participants with Schizophrenia						
Dependent Variable	Step	Predictor	$R^2_{Incremental}$	R^2_{total}	t	p
Trial-by-Trial Confidence	1	Accuracy Judgments	.33	.33	6.96	<.001
Dependent Variable	Step	Predictor				
Global Performance Judgments	1	Trial × Trial Confidence	.26	.26	6.25	<.001
	2	Accuracy Judgments	.20	.46	5.79	<.001
Participants with Bipolar Illness						
Dependent Variable	Step	Predictor	$R^2_{Incremental}$	R^2_{total}	t	p
Trial-by-Trial Confidence	1	WCST Performance	.24	.24	2.93	.005
	2	Accuracy Judgments	.05	.29	2.11	.038
Dependent Variable	Step	Predictor				
Global Performance Judgments	1	WCST Performance	.57	.57	7.51	<.001
	2	Trial × Trial Confidence	.04	.61	3.12	.003
	3	Accuracy Judgments	.02	.63	2.05	.045

Note. Higher levels of self-reported trial-by-trial accuracy judgments in bipolar participants were negatively correlated with global judgments of performance.