Schizophrenia Research: Cognition 32 (2023) 100278

Contents lists available at ScienceDirect

Schizophrenia Research: Cognition

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

Research Paper

Dynamics of task-based confidence in schizophrenia using seasonal decomposition approach

Varsha D. Badal [a,b], Colin A. Depp [a,b,c,*], Amy E. Pinkham [d], Philip D. Harvey [e,f]

[a] *Department of Psychiatry, University of California San Diego, San Diego, CA, USA*
[b] *Sam and Rose Stein Institute for Research on Aging, University of California San Diego, San Diego, CA, USA*
[c] *VA San Diego Healthcare System, La Jolla, CA, USA*
[d] *Department of Psychology, School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, TX, USA*
[e] *Department of Psychiatry and Behavioral Sciences, University of Miami, Miller School of Medicine, Miami, FL, USA*
[f] *Research Service, Miami VA Healthcare System, USA*

A R T I C L E I N F O

Keywords:
Time series
Psychosis
Metacognition
Executive function
Social cognition

A B S T R A C T

Objective: Introspective Accuracy (IA) is a metacognitive construct that refers to alignment of self-generated accuracy judgments, confidence, and objective information regarding performance. IA not only refers to accuracy and confidence during tasks, but also predicts functional outcomes. The consistency and magnitude of IA deficits suggest a sustained disconnect between self-assessments and actual performance. The cognitive origins of IA are unclear and are not simply due to poor performance. We tried to capture task and diagnosis-related differences through examining confidence as a timeseries.
Method: This relatively large sample (N = 171; Bipolar = 71, Schizophrenia = 100) study used item by item confidence judgments for tasks including the Wisconsin Card Sorting Task (WCST) and the Emotion Recognition task (ER-40). Using a seasonal decomposition approach and AutoRegressive, Integrative and Moving Averages (ARIMA) time-series analyses we tested for the presence of randomness and perseveration.
Results: For the WCST, comparisons across participants with schizophrenia and bipolar disorder found similar trends and residuals, thus excluding perseverative or random responding. However, seasonal components were weaker in participants with schizophrenia, reflecting a reduced impact of feedback on confidence. In contrast, for the ER40, which does not require identification of a sustained construct, seasonal, trend, and residual analyses were highly comparable.
Conclusion: Seasonal analysis revealed that confidence judgments in participants with schizophrenia on tasks requiring responses to feedback reflected diminished incorporation of external information, not random or preservative responding. These analyses highlight how time series analyses can specify potential faulty processes for future intervention.

1. Introduction

Self-assessment is a major challenge in schizophrenia across multiple content domains, including symptoms, cognition, functional abilities, and real-world functioning. One construct that links these domains is Introspective Accuracy (IA), the ability to generate self assessments that are convergent with objective evidence. IA is functionally relevant, as multiple studies, using a variety of methods to operationalize IA, have linked IA challenges with poorer real-world functioning in severe mental illness (Gould et al., 2015, Silberstein and Harvey, 2019, Durand et al., 2021, Pinkham et al., 2018, Harvey et al., 2019). Since IA impairments share features across domains, including self-assessments of functioning, cognition, and social cognition (Durand et al., 2021; Silberstein and Harvey, 2019) and measurement approaches, it offers promise as a potential target for cognitive training (Reiter et al., 2021).

IA impairments appear greater in psychotic disorders (Tercero et al., 2021; Durand et al., 2021) where global self-assessments of competence are significantly more disconnected from actual performance compared to bipolar disorder. The dominant directional bias is consistent with over estimation of ability and overconfidence (Perez et al., 2020) and

* Corresponding author at: Stein Institute for Research on Aging, Department of Psychiatry (0664), University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0664, USA.
E-mail address: cdepp@ucsd.edu (C.A. Depp).

https://doi.org/10.1016/j.scog.2023.100278
Received 11 January 2023; Received in revised form 12 January 2023; Accepted 16 January 2023

overestimation occurs in participants with schizophrenia across the entire range of performance (Jones et al., 2019), suggesting that limited ability is not the only driver.

Confidence judgments on cognitive tasks can offer insights into the dynamics of metacognitive processes. Based on the definition of confidence (Dunlosky and Metcalfe 2009), self-reported confidence can be considered an expression of the quality of assessment of one's current state of knowledge. Confidence also acts as a proxy for knowledge when there is incomplete information available to solve the task, and in some ways, it can be seen as a measure of how much trust was placed in the available evidence (Ptasczynski et al., 2021). An initial investigation into metacognition using machine learning had revealed confidence responses aided in discriminating between the schizophrenia and healthy control groups beyond what was feasible using only the performance variables (Badal et al., 2021), possibly in a complex nonlinear manner. Owing to the opaque nature of methods, no further insights were available. The role of IA in context of task structure and performance was considered the key to this understanding. Confidence was shown to play a role in performance, not only in the moment, but also in the future (Pinkham et al., 2018; Badal et al., 2022).

Successful performance on some tasks requires a series of correct responses using the same criteria; these tasks should display an accrual of confidence based on a time series of correct response. Other tasks would not be expected to see accrual of confidence across responses, particularly if there is little information available about how to solve the next item based on performance on the current item. The Wisconsin Card Sorting Task (WCST) (Koren et al., 2006) is a useful paradigm for assessing accrued confidence processes. The task was designed to assess development of knowledge through experience and cognitive flexibility (Heaton, 1981), requiring both the ability to retain successful strategies, and quickly readapt when feedback is negative (Feng et al., 2020). In contrast, the Emotion Recognition task 40-item version (ER40), asks participants to identify the emotion shown on various faces (Gur et al., 2002), with the series of emotional expressions unrelated to each other and variable in their difficulty. Our modified version of the task includes confidence and IA measures (Pinkham et al., 2018; Badal et al., 2022). This task has little item-to-item coherence or carryover, with no need for incorporation of feedback for fine-tuning subsequent responses. The ER40 hence forms a comparative paradigm for the assessment of dynamics of confidence judgments against the WCST, with the critical shared features being a sustained time series of stimuli and responses, with performance, accuracy judgments, and confidence in those judgments collected in both paradigms.

There are several possible approaches to evaluating tasks as a time series while addressing the differences between feedback dependent and independent problem solving. Our previous study used a network approach to understand the dynamics of confidence (Badal et al., 2022), over-confidence foreshadowed poor future performance in participants with schizophrenia compared to other groups, suggesting poorer incorporation of externally generated evidence. However, some critical performance features were not considered: 1) We did not address the possibility of random responses defined by markedly different or excessively variable trajectories of confidence or disproportionate residuals, 2) we did not address the presence of preservative behavior defined by a constant trajectory or minimal variance, and 3) our network approach could not highlight characteristic differences between the tasks; the WCST demands discovery of sort criteria and sustaining that construct over trials, which should be very evident in seasonal decomposition approach. These issues are important because incomplete information on presence of randomness and perseveration limits our understanding of task demands and cognitive deficits in schizophrenia. These types of responses may be less amenable to a rehabilitation intervention. This approach was lacking in our previous study (Badal et al., 2022), which dealt with the relationship among task performance, confidence and performance judgments.

Continuous runs (epochs) of task items that have a similar solution (and hence similar anticipated feedback based on correct responses), can be expected to be internalized as a "trend" or "season" in a metacognitive variable such as confidence. For example, when an individual gets 5 correct responses in a row with feedback provided, they should be more likely to also anticipate that the 6th response will be correct as well. But such scrutiny is not possible with network analysis owing to stationarity assumptions. As sort criteria are maintained for a fixed duration, and each criterion (i.e., category), and the correct sorts associated with it, can be considered as a season, confidence measures on the entire WCST can be seen as a sequence of sort-defined seasons that can be subject to appropriate seasonal analysis (Cleveland et al., 1990). Autoregression, Integrated with Moving Average (ARIMA) allows us to capture the degrees of polynomials that best describe the trend and the seasonal components.

Motivated by this reasoning, we hypothesize that the poorer incorporation of feedback among people with schizophrenia compared to the Bipolar group will be manifested in the trajectory of confidence within seasons (specifically, during a sequence of correct sorts within a specific category, confidence will increase in participants with bipolar disorder in sync with feedback, but not for participants with schizophrenia). We also hypothesize that in tasks where there is little item-to-item commonality, such as in ER40, differences between participants with bipolar disorder and participants with schizophrenia would be less pronounced.

2. Method

2.1. Participants

Participants were outpatients recruited from three universities: 1) University of California San Diego (UCSD), and associated Outpatient Psychiatric Services clinic and the San Diego VA Medical Center 2) The University of Texas at Dallas (UTD) and Metrocare Services in Dallas County, Texas and 3) The University of Miami (UM) and Jackson Memorial Hospital-University of Miami Medical Center and the Miami VA Medical Center. This analysis includes patients with diagnoses of schizophrenia or schizoaffective disorder ($n = 100$), bipolar I disorder ($n = 71$).

Participant inclusion criteria comprised of 1) a DSM-V diagnosis of Schizophrenia (and Schizoaffective disorder) or Bipolar I disorder (with or without psychotic features) 2) ages between 18 and 65 years, 3) ability to communicate in English, 4) outpatient, 5) stable medications for at least 6 weeks, and 6) willingness to provide a high contact informant with no prior psychiatric diagnosis. Exclusion criteria for the study were: 1) a history of or current medical or neurological disorders that might affect brain functioning (e.g. stroke, untreated seizures, history of loss of consciousness >15 min), 2) low estimated verbal IQ (i.e., a standard score <70 on the Wide Range Achievement Test 4 Reading test (Wilkinson and Robertson, 2006) (or pervasive developmental disorder according to the DSM-V criteria, 3) substance use disorder (excluding tobacco and cannabis) in the past six months and 4) any visual or hearing impairments that might interfere with assessments. Participants were also excluded if they had been hospitalized within the past six weeks. All participants were required to provide signed informed consent. The studies were approved by institutional review boards at each of the corresponding sites.

2.2. Measures

2.2.1. The Modified Wisconsin Card Sorting Task (WCST)

The WCST is a standard neuropsychological test of cognitive flexibility (Heaton, 1981). A modified version of the WCST task, called the Metacognitive WCST (Tercero et al., 2021), was administered in this study. This version of the test presents participants with a sequence of 64 cards, in which the participants are instructed to sort as in the standard WCST without being given any information about sorting criteria, with the standard color-form-number sequence of categories. The sort criteria changes to the next category after 10 consecutive correct responses. For each item in the task three responses from the participant were recorded: 1) the sort, 2) the participant's judgment on response accuracy (Did you get it correct? – Yes/No), and 3) the participant's confidence in the correctness of their judgment about their accuracy (on a 5-point scale from 0 %–100 % confident). After having recorded the above responses, the feedback regarding actual accuracy was provided to the participant.

2.2.2. The Modified Emotion Recognition (ER40)

Is a modified version of the long-standing 40-item task (Gur et al., 2002) which requires identification of emotions of varying strengths depicted in the photographs of human faces (Badal et al., 2022). The task covers four basic emotions (happiness, sadness, anger, and fear) and neutral expressions, each in equal proportions, which were presented, one face at a time. Participants were required to identify the emotion depicted in the face, which varied in the strength of the emotion and in difficulty. This was followed by self-assessment of accuracy judgment (Did you get it correct? – Yes/No) and confidence (on a 5-point scale from 0 % to 100 % confident). The participant was then provided feedback on accuracy of identified emotion.

2.3. Analysis

2.3.1. Seasonal decomposition

The study utilizes seasonal decomposition which is based upon Wold's theorem (Anderson, 1971) which proposes that a time series can be seen as a sum of deterministic and stochastic components. The classical method uses a moving average method to estimate the trend. Once the trend is established, the data is detrended by utilizing the trend and then seasonal effects and residuals are calculated (Hyndman and Athanasopoulos, 2018). The decomposition can be written as:

$$y_t = T_t + S_t + R_t \tag{1}$$

where y_t is the observed signal, T_t is the trend component, S_t is the seasonal component and R_t is the residual. We have used the python implementation of classical method in *statsmodels* module (Seabold and Perktold, 2010). The cycle length for seasonal decomposition was set to 10 to correspond to the length of correct responses when new sort criteria were introduced in the WCST. In the ER-40, there was no such seasonality, and the 10-item sequences were used for comparison purposes.

In order to assess the absolute quantitative measure of strength of seasonal components we used the Root Mean Squared (RMS) value of the signal which is often used as a measure of energy or effective value in signal processing (Daintith, 2009). A measure of trend and seasonal strength relative to its residual or error term can also be evaluated (Wang et al., 2006).

2.3.2. ARIMA

ARIMA models have long been used to analyze timeseries and in forecasting applications (Shumway and Stoffer, 2017; Ho and Xie, 1998; Hyndman and Athanasopoulos, 2015). This acronym captures three key aspects of the model, they are: 1) The *Autoregressive (AR)* component captures the contribution to timeseries from its previous values, 2) the *Integrated (I)* component assesses the stationarity through differencing, and 3) the *Moving Average (MA)* utilizes moving average to establish trend.

3. Results

3.1. Participants

Participant ages ranged between 19 and 64 years, and the proportion of women was significantly higher in among the participants with bipolar disorder group (70.4 % vs 48.0 %, $\chi^2 = 7.64$ p = .006, Table 1).

Table 1
Demographic details of participants.

	Bipolar I disorder (N = 71)	Schizophrenia (N = 100)	Schizophrenia vs. bipolar I disorder		
			t-Test χ^2	p-Value	Cohen's d
Age — mean (SD); range	39.2 (11.8); 19–64	41.9 (10.5); 20–64	−1.58	.116	0.25
Gender — female (%)	50 (70.4 %)	48 (48.0 %)	7.64	**.006**	−0.46
Education in years — mean (SD)	14.3 (2.6)	12.5 (2.3)	4.61	**<.001**	−0.72
Single (%)	35 (49.3 %)	67 (67.0 %)	4.70	**.030**	
Race					
Caucasian (%)	54.9 %	32.0 %	8.07	**.004**	
African American (%)	26.8 %	54.0 %	11.50	**.001**	
Other (%)	18.3 %	14.0 %	0.30	.583	
Ethnicity					
Hispanic (%)	29.6 %	23.0 %	0.63	.428	
Vocation (some overlap between categories)					
Unemployed	56.3 %	71.0 %	3.30	.069	
Clinical measures					
WRAT-3 — mean (SD)	101.9 (11.6)	95.6 (11.8)	3.46	**.001**	−0.54
PANSS positive symptoms — mean (SD)	13.4 (4.9)	17.2 (4.6)	−5.24	**.000**	0.81
PANSS negative symptoms — mean (SD)	10.8 (2.6)	13.5 (4.3)	−4.87	**.000**	0.76
MADRS total — mean (SD)	12.9 (11.4); 0–37	9.3 (10.1); 0–38	2.14	**.034**	−0.33
ER40 — mean (SD); range					
Correct faces (out of 40)	26.1 (12.0); 0–40	25.1 (11.7); 0–40	0.71	.479	−0.15
Confidence – (100-point scale)	78.1 (21.4); 0–100	76.8 (22.4); 0–100	0.38	.707	−0.06
Estimated correct faces (out of 40)	35.1 (9.9); 0–40	33.5 (10.7); 0–40	1.01	.315	−0.16
Correct and estimated faces match	31.3 (5.5); 0–40	29.8 (6.1); 0–40	1.68	.094	−0.26
WCST — mean (SD); range					
Correct sorts (out of 64)	36.2 (13.1); 0–64	30.4 (11.9); 0–64	3.00	**.003**	−0.47
Confidence-100-point scale	73.3 (22.3); 0–100	73.5 (20.6); 0–100	−0.08	.940	0.01
Estimated correct sort (out of 64)	51.1 (18.6); 0–64	48.6 (18.0); 0–64	0.90	.370	−0.14
Correct and estimated sort match	39.6 (10.4); 0–64	34.4 (9.9); 0–64	3.29	**.001**	−0.51

WRAT-3: Wide Range Achievement Test 3; PANSS: Positive and Negative Syndrome Scale; MADRS: Montgomery–Åsberg Depression Rating Scale; ER40: Penn Emotion recognition task; WCST: Wisconsin Card Sorting Test.
Significant p-values are in bold font.

The socio-demographic differences between groups were not unexpected and were in line with previous studies (Tercero et al., 2021; Jones et al., 2021; Harvey et al., 2021; Durand et al., 2021), and along generally accepted dimensions of race, employment, marital status, clinical scales and cognitive skills. The participants with bipolar I disorder had higher educational attainment as measured by years of education completed (14.3 vs. 12.5, $\chi^2 = 4.61$ $p < .001$) and were less likely to be single (49.3 % vs 67.0 %, $\chi^2 = 4.70$ $p = .030$). Distribution of races was also significantly different with more African American individuals in the participants with schizophrenia group.

Secondary analyses were performed on subsets that excluded participants who did not correctly complete a single WCST category (defined by never identifying any sort criterion with 10 consecutive correct responses), to eliminate outliers including those who may not have understood the task requirements. These reduced subsets comprised of participants with schizophrenia (n = 57; 57 %) and participants with bipolar disorder (n = 54; 76 %).

3.2. Seasonal decomposition of averaged confidence for WCST task

The seasonal decomposition of average item measure of confidence for the WCST task is shown in (Fig. 1). The top panel shows the confidence for each item (a total of 64 items) averaged for each of the groups that is decomposed into a trend in the second panel and a seasonal component in the third. The fourth panel shows the residual values.

3.2.1. Observed confidence

In the beginning (task items 1–4, labelled Gap 0), both groups display declining confidence as the chief task is to identify that the correct sort criteria is color and not number or shape. Thereafter, the participants with schizophrenia and participants with bipolar disorder groups diverged with participants with bipolar disorder showing increasing

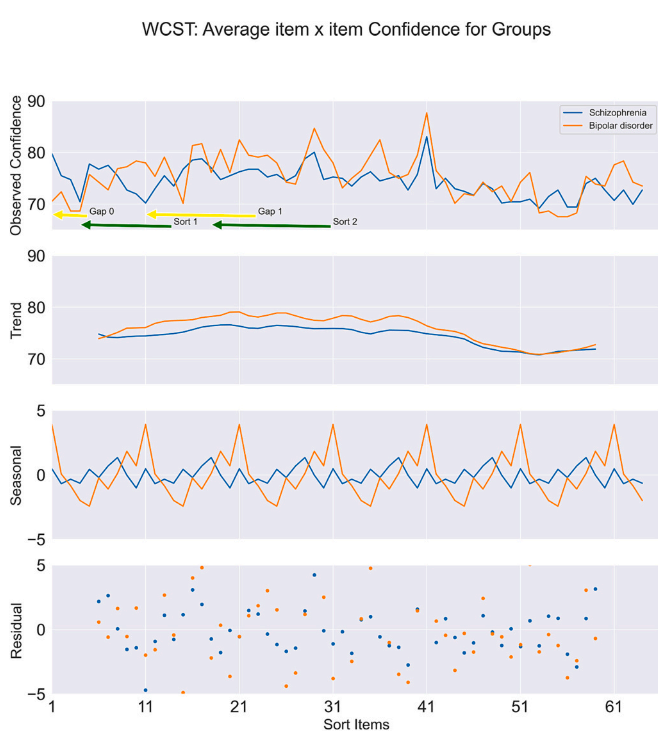

Fig. 1. Seasonal decomposition of item-by-item confidence in the WCST task averaged across the diagnostic categories (participants with schizophrenia and schizoaffective disorders, participants with bipolar I disorder). The top panel displays the constructed time series. The second panel displays the trend component, the third panel displays the seasonal component, and the fourth panel displays the residuals.

confidence as participants on average identify the correct sort criteria (items 4–12, labelled Sort1). This is followed by a sudden drop in the average confidence centered around items 13–14 as the sort-criteria changes without the prior knowledge of the participant. The graph for participants with schizophrenia shows a similar but less well-defined phase, as confidence does not rise as much as it does in participants with bipolar disorder.

This is followed by regaining of confidence in both participants with schizophrenia and participants with bipolar disorder for the sort criteria based on number, however many participants take longer to identify and increasingly fail to switch to subsequent sorting categories (items 11–24 Gap 1, and Sort 2 region).

3.2.2. Trend component

The trend components for the two groups represent a smoothed version of the experience described above, and they have similar starting and ending values, however, the component is generally higher for the participants with bipolar disorder compared to the participants with schizophrenia in the starting and middle phase of the 64-item task. We repeated this analysis with only those participants who got at least one category correct. In this subgroup, the trend component for participants with bipolar disorder was higher throughout compared to participants with schizophrenia (0.58 vs. 0.47).

3.2.3. Seasonal component

The seasonal components show greater seasonal amplitude for the participants with bipolar disorder compared to participants with schizophrenia (Fig. 1). The absolute strength as measured by the RMS components are about twice as strong for the participants with bipolar disorder compared to participants with schizophrenia (1.79 vs. 0.68) as were the relative measures (0.26 vs. 0.11) (Table 2), suggesting participants with bipolar disorder display greater sensitivity to feedback than participants with schizophrenia. The absolute component was also stronger for participants with bipolar disorder for the subset comprising participants who got at least one category correctly identified by a continuous run of 10 correct (1.5 vs.0.79) however the gap narrowed. Even though participants with bipolar disorder continued to display greater sensitivity than participants with schizophrenia, narrowing of the gap between relative components was also observed in this subsample (0.24 vs. 0.15), suggesting groups were less heterogenous when participants who identified no category correctly were removed. A sizeable fraction (42.4 %) of participants with schizophrenia did not complete one category. When compared to that of the whole group, the relative trend (0.39 vs 0.48) was moderately weaker, and the seasonal components (0.10 vs 0.11) were comparable. A much smaller fraction (19.4 %) of the bipolar group failed to complete one category, and the relative trend (0.01 vs. 0.46) was much weaker but seasonality (0.28 vs. 0.26) remained intact. This suggests that negative performance feedback in people with schizophrenia had less influence on confidence trends than in people with bipolar disorder.

3.2.4. Residual component

The residual components increased after the first 13 or so items for both groups which explains greater variation as participants manifest increasing variance in their progress toward identification of subsequent categories as the task progresses.

3.3. Seasonal decomposition of averaged confidence for ER40 task

The seasonal decomposition of confidence in the ER40 task is shown in Fig. 2. As in the previous figure, the top panel shows the confidence for each item (a total of 40 items) averaged for each of the groups that is decomposed into a trend in the second panel and a seasonal component in the third. The fourth panel shows the residual values.

Table 2
Measures of strength of trend and seasonal components.

	WCST (all participants)			WCST (at least one category sorted correctly and completely)			ER40 (all participants)		
	T_R	$S_{Effective}$	S_R	T_R	$S_{Effective}$	S_R	T_R	$S_{Effective}$	S_R
Schizophrenia	0.48	0.68	0.11	0.47	0.79	0.15	0.29	1.82	0.34
Bipolar I disorder	0.41	1.79	0.26	0.58	1.50	0.24	0.20	2.74	0.32

T_R: Strength of trend relative to residual term.
S_R: Strength of season relative to residual term.
$S_{Effective}$: RMS.

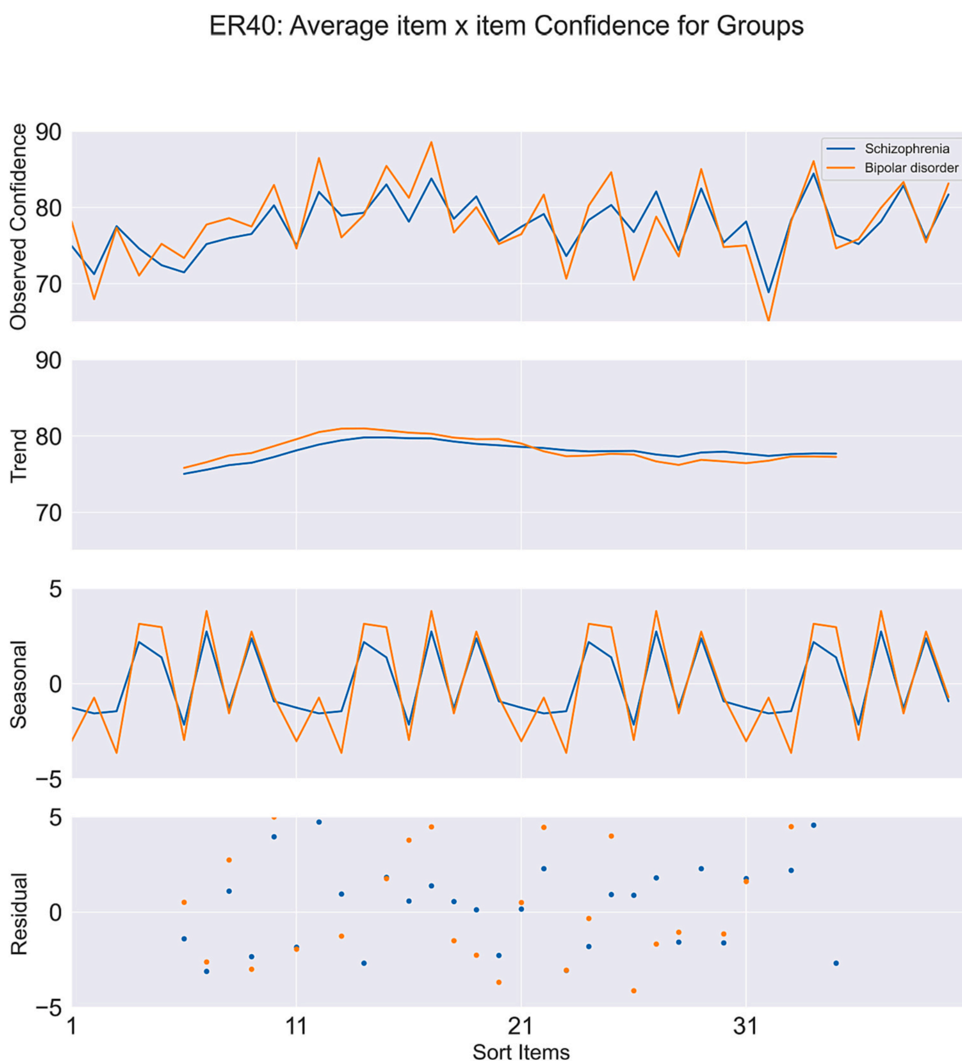

Fig. 2. Seasonal decomposition of item-by-item confidence in the ER40 task averaged across the diagnostic categories (participants with schizophrenia and schizoaffective disorders, participants with bipolar I disorder). The top panel displays the constructed time series. The second panel displays the trend component, the third panel displays the seasonal component, and the fourth panel displays the residuals.

3.3.1. *Observed confidence*

The item-to-item changes in confidence are much larger in the ER40 task compared to the WCST task, possibly due to greater variation in task difficulty from item to item and no carry-over of problem solving across stimuli. Confidence appears very similar for the two groups.

3.3.2. *Trend component*

The trend components for the two groups are also quite similar in the starting, middle, and ending phase of the 40-item task. The average confidence is lower in the beginning of the task, which soon increases and then declines slightly.

3.3.3. *Seasonal component*

The decomposition did not show a distinct seasonal building pattern, likely because of the lack of relatedness of the sequential items and the item-by-item fluctuation overshadows a season long (10 item) cycle. Although the RMS value was greater for participants with bipolar disorder, the relative value was greater for participants with schizophrenia (Table 2).

3.3.4. *Residual component*

The residues were generally higher than those in the WCST task. The middle portion of the task (items 15–30) displayed less variation than the beginning and the end.

3.4. ARIMA analysis of WCST

ARIMA captures the order of three components of the time series: Auto Regressive (AR) as the first, the Integrative component (I) as the second and the Moving Average (MA) as the third. Of the three, the auto regressive component, which is the first number in the triple is expected to capture inertia in confidence judgments. Best fitting ARIMA models (see Appendix for detailed results) suggest the presence of an autoregressive component in participants with schizophrenia (ARIMA (1,1,1) implies AR = 1, I = 1, MA = 1) but not in participants with bipolar disorder (ARIMA (0,1,2) implies AR = 0, I = 1, MA = 2). Integrative and moving average components (the second and the third components of the triple) were present in both the groups. The presence of (I) term in both groups suggests confidence is non-stationary over the test window (64 items in WCST, and 40 items in ER40), and it may have ended at a different level than where it started. The moving average terms suggests the change in participants with schizophrenia it is linear (1st order) vs. quadratic (2nd order) in participants with bipolar disorder, suggesting quicker change in participants with bipolar disorder.

4. Discussion

In this study we performed a univariate analysis of averaged item-by-item confidence for the schizophrenia and bipolar disorder groups in this novel manner. Our findings suggest that item-to-item accrual of confidence is captured very well by the seasonal components when season length was set to correspond to the underlying length of 10 correct items corresponding to a single WCST category. For the WCST, since the seasonal component for participants with bipolar disorder as measured by the RMS ($S_{Effective}$) and the relative (S_R) is about twice as strong as in the participants with schizophrenia, participants with bipolar disorder display cognitive flexibility through appropriate modulation of confidence in line with momentary sorting success, whereas participants with schizophrenia show a relatively greater inertia. It is worthwhile to note that this inertia (e.g., responses of 1, 1, 2, 1, 1, etc.) is not perseverative, by which we mean exact repetition of past responses (e.g., responses of all 1 s). Smaller updates to confidence in participants with schizophrenia also show up in presence of AR (autoregressive) component of ARIMA which captures the degree to which responses can be predicted from the previous value. Very similar trend profiles, and strong relative trend components coupled with presence of AR component in participants with schizophrenia supports the notion that they are not responding randomly. The findings were consistent even among the subgroups that completed at least one sort-criterion correctly.

Studies do indeed suggest that confidence has a carryover or accrual effects across trials, a part of which may be general (Mazancieux et al., 2020) and a part that is domain specific (Fitzgerald et al., 2017). The domain specific carryover is often the greatest when classifications are repeated (Kantner et al., 2019).

In our previous study (Badal et al., 2022) we related hyperfocusing theory (Luck et al., 2019) with confidence leakage (Shekhar and Rahnev, 2021). According to the hyperfocusing theory, participants with schizophrenia excessively focus upon evidence consistent with their working memory contents while ignoring external cues, and the confidence leakage theory suggests that past confidence judgments in participants with schizophrenia influence future judgments even when they should not. The overall effect is of participants with schizophrenia misplacing their efforts while poorly incorporating evidence. This study supports the idea that the phenomena of item-to-item confidence leakage is at work in participants with schizophrenia, whereby confidence shows inertia and changes in relatively smaller steps from item-to-item despite accuracy feedback, whereas participants with bipolar disorder shows a more fluid confidence response that accrues with positive feedback and subsequently notably dissipates with unexpected errors and consequential uncertainty as to the next correct response.

The weaker trend component in ER40 is expected as there is no underlying trend in unrelated items sequentially presented to the participant. The stronger relative seasonality and absolute seasonality for the ER40 task is likely function of the individual difficulty of the items, as there is more variance in the confidence judgments allowing for greater area under the swings that goes into these calculations. In absolute terms participants with bipolar disorder do show greater fluidity of confidence as measured using RMS ($S_{Effective}$) than the participants with schizophrenia, suggesting greater sensitivity to feedback. As noted, before, the lower absolute measure of seasonality for the participants with schizophrenia is likely due to confidence leakage that prevents big changes in confidence.

There are some limitations of this study as it relies upon data from outpatients only and did not include healthy comparators. It can be suggested that other factors such as severity of psychotic symptoms may influence confidence but were not acknowledged by the model, however our previous study found no correlations between PANSS positive and confidence or accuracy (Sabbag et al., 2012; Harvey et al., 2017). Univariate timeseries analysis techniques such as ARIMA and seasonal decomposition have little ability to estimate covariate effects.

In summary, for the WCST task, we found evidence that participants with bipolar disorder show a seasonal component in confidence that is tied to underlying sort criteria and their intrinsic variation over time in the task. Our results support the idea that people with schizophrenia do not respond randomly or preservatively, rather their confidence seems to be adjusted sub-optimally possibly leading to inaccurate decision making. Evidence from a study on healthy controls suggests that such shortfalls are modifiable through adaptive training by targeting confidence which drives future shifts in metacognitive efficiency (Carpenter et al., 2019). Interventions that improve the accuracy of confidence judgments in participants with schizophrenia could impact metacognition and ultimately functional outcome.

Funding support

This work was supported by the National Institute of Mental Health (R01MH116902 to C.A.D.; R0MH112620 to A.E.P.). Salary for V.D.B. was supported by National Institute of Mental Health (NIMH) T32 Geriatric Mental Health Program MH019934 to Dilip V. Jeste.

Transparency and openness

The data is available through National data Archive maintained by NIH (https://nda.nih.gov/faq.html). Seasonal decomposition and ARIMA was performed using python libraries and the online source repositories have been provided for reference. All parameters or customizations to analyze the data have been described in the manuscript in sufficient detail.

CRediT authorship contribution statement

VDB: Designed and implemented the timeseries based approaches, performed data analysis, and interpreted results, prepared the manuscript, and contributed to drafts of the manuscript.
CAD: Designed and oversaw the study, proposed the research questions, helped with data analysis and interpretation, provided feedback, and contributed to drafts of the manuscript.
AEP: Designed and oversaw the study, provided feedback, and contributed to drafts of the manuscript.
PDH: Designed and oversaw the study, proposed the research questions, provided feedback, and contributed to drafts of the manuscript.

All authors have reviewed, edited the paper, and approve the final version.

Conflict of interest

P.D.H. has received consulting fees or travel reimbursements from Alkermes, Bio Excel, Boehringer Ingelheim, Karuna Pharma, Minerva Pharma, SK Pharma, and Sunovion Pharma during the past year. He receives royalties from the Brief Assessment of Cognition in Schizophrenia (Owned by Verasci, Inc. and contained in the MCCB). He is chief scientific officer of i-Function, Inc. R.C.M. is a co-founder of KeyWise, Inc. and a consultant for NeuroUX.

For the remaining authors, no conflicts of interest were declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scog.2023.100278.

References

Anderson, T.W., 1971. The statistical analysis of time series. Wiley, New York.

Badal, V.D., Depp, C.A., Harvey, P.D., Ackerman, R.A., Moore, R.C., Pinkham, A.E., 2022. Confidence, accuracy judgments and feedback in schizophrenia and bipolar disorder: a time series network analysis. Psychol. Med. 1–10.

Badal, V.D., Depp, C.A., Hitchcock, P.F., Penn, D.L., Harvey, P.D., Pinkham, A.E., 2021. Computational methods for integrative evaluation of confidence, accuracy, and reaction time in facial affect recognition in schizophrenia. Schizophr. Res.: Cogn. 25, 100196.

Carpenter, J., Sherman, M.T., Kievit, R.A., Seth, A.K., Lau, H., Fleming, S.M., 2019. Domain-general enhancements of metacognitive ability through adaptive training. J. Exp. Psychol. Gen. 148, 51.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. STL: a seasonal-trend decomposition. J. Off. Stat. 6, 3–73.

Daintith, J., 2009. A Dictionary of Physics. Oxford University Press, Oxford; New York.

Dunlosky, J., Metcalfe, J., 2008. Metacognition, Sage Publications..

Durand, D., Strassnig, M.T., Moore, R.C., Depp, C.A., Ackerman, R.A., Pinkham, A.E., Harvey, P.D., 2021. Self-reported social functioning and social cognition in schizophrenia and bipolar disorder: using ecological momentary assessment to identify the origin of bias. Schizophr. Res. 230, 17–23.

Feng, X., Perceval, G.J., Feng, W., Feng, C., 2020. High cognitive flexibility learners perform better in probabilistic rule learning. Front. Psychol. 11, 415.

Fitzgerald, L.M., Arvaneh, M., Dockree, P.M., 2017. Domain-specific and domain-general processes underlying metacognitive judgments. Conscious. Cogn. 49, 264–277.

Gould, F., McGuire, L.S., Durand, D., Sabbag, S., Larrauri, C., Patterson, T.L., Twamley, E.W., Harvey, P.D., 2015. Self-assessment in schizophrenia: accuracy of evaluation of cognition and everyday functioning. Neuropsychology 29, 675–682.

Gur, R.C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., Turner, T., Bajcsy, R., Posner, A., Gur, R.E., 2002. A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. J. Neurosci. Methods 115, 137–143.

Harvey, P.D., Miller, M.L., Moore, R.C., Depp, C.A., Parrish, E.M., Pinkham, A.E., 2021. Capturing clinical symptoms with ecological momentary assessment: convergence of momentary reports of psychotic and mood symptoms with diagnoses and standard clinical assessments. Innov. Clin. Neurosci. 18, 24.

Harvey, P.D., Strassnig, M.T., Silberstein, J., 2019. Prediction of disability in schizophrenia: symptoms, cognition, and self-assessment. J. Exp. Psychopathol. 10, 2043808719865693.

Harvey, P.D., Twamley, E.W., Pinkham, A.E., Depp, C.A., Patterson, T.L., 2017. Depression in schizophrenia: associations with cognition, functional capacity, everyday functioning, and self-assessment. Schizophr. Bull. 43, 575–582.

Heaton, R.K., 1981. Wisconsin card sorting test manual. Psychological assessment resources.

Ho, S.L., Xie, M., 1998. The use of ARIMA models for reliability forecasting and analysis. Comput. Ind. Eng. 35, 213–216.

Hyndman, R.J., Athanasopoulos, G., 2015. 8.9 Seasonal ARIMA models. In: Forecasting: Principles and Practice. oTexts. Retrieved, 19.

Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: principles and practice. In: OTexts.

Jones, M.T., Deckler, E., Laurrari, C., Jarskog, L.F., Penn, D.L., Pinkham, A.E., Harvey, P. D., 2019. Confidence, performance, and accuracy of self-assessment of social cognition: A comparison of schizophrenia patients and healthy controls. Schizophrenia Research: Cognition.*19*, p.100133.

Jones, S.E., Moore, R.C., Depp, C.A., Ackerman, R.A., Pinkham, A.E., Harvey, P.D., 2021. Daily ecological momentary assessments of happy and sad moods in people with schizophrenia and bipolar disorders: what do participants who are never sad think about their activities and abilities? Schizophr. Res. Cogn. 26, 100202.

Kantner, J., Solinger, L.A., Grybinas, D., Dobbins, I.G., 2019. Confidence carryover during interleaved memory and perception judgments. Mem. Cogn. 47, 195–211.

Koren, D., Seidman, L.J., Goldsmith, M., Harvey, P.D., 2006. Real-world cognitive–and metacognitive–dysfunction in schizophrenia: a new approach for measuring (and remediating) more "right stuff". Schizophr. Bull. 32, 310–326.

Luck, S.J., Hahn, B., Leonard, C.J., Gold, J.M., 2019. The hyperfocusing hypothesis: a new account of cognitive dysfunction in schizophrenia. Schizophr. Bull. 45, 991–1000.

Mazancieux, A., Fleming, S.M., Souchay, C., Moulin, C.J., 2020. Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. J. Exp. Psychol. Gen. 149, 1788.

Perez, M.M., Tercero, B.A., Penn, D.L., Pinkham, A.E., Harvey, P.D., 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.

Pinkham, A.E., Harvey, P.D., Penn, D.L., 2018a. Social cognition psychometric evaluation: results of the final validation study. Schizophr. Bull. 44, 737–748.

Ptasczynski, L.E., Steinecker, I., Sterzer, P., Guggenmos, M., 2021. The value of confidence: Confidence prediction errors drive value-based learning in the absence of external feedback.

Reiter, A.M.F., Atiya, N.A.A., Berwian, I.M., Huys, Q.J.M., 2021. Neuro-cognitive processes as mediators of psychological treatment effects. Curr. Opin. Behav. Sci. 38, 103–109.

Sabbag, S., Twamley, E.W., Vella, L., Heaton, R.K., Patterson, T.L., Harvey, P.D., 2012. Predictors of the accuracy of self assessment of everyday functioning in people with schizophrenia. Schizophr. Res. 137, 190–195.

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python. In: Proceedings of the 9th Python in Science Conference, p. 61. Austin, TX.

Shekhar, M., Rahnev, D., 2021. Sources of metacognitive inefficiency. Trends Cogn. Sci. 25, 12–23.

Shumway, R.H., Stoffer, D.S., 2017. ARIMA Models. Time Series Analysis and Its Applications: With R Examples. Springer International Publishing, Cham.

Silberstein, J., Harvey, P.D., 2019. Impaired introspective accuracy in schizophrenia: an independent predictor of functional outcomes. Cogn. Neuropsychiatry 24, 28–39.

Tercero, B.A., Perez, M.M., Mohsin, N., Moore, R.C., Depp, C.A., Ackerman, R.A., Pinkham, A.E., Harvey, P.D., 2021. Using a meta-cognitive Wisconsin card sorting test to measure introspective accuracy and biases in schizophrenia and bipolar disorder. J. Psychiatr. Res. 140, 436–442.

Wang, X., Smith, K., Hyndman, R., 2006. Characteristic-based clustering for time series data. Data Min. Knowl. Disc. 13, 335–364.

Wilkinson, G.S., Robertson, G.J., 2006. Wide Range Achievement Test (WRAT4). Psychological Assessment Resources, Lutz, FL.