GPT-3 reveals selective insensitivity to global *vs.* local linguistic context in speech produced by treatment-naïve patients with positive thought disorder

Victoria Sharpe[1], Michael Mackinley[2,], Samer Nour Eddine[1,3], Lin Wang[1,3], Lena Palaniyappan[4,5], and Gina R. Kuperberg[1,3]

[1]Department of Psychology, Tufts University,

[2]Lawson Health Research Institute

[3]Massachusetts General Hospital, Harvard Medical School

[4]Robarts Research Institute, Western University, Ontario

[5]Douglas Mental Health University Institute, McGill University

Corresponding Author:

Victoria Sharpe, PhD

Tufts University, Department of Psychology

490 Boston Avenue

Medford, MA 02155

e-mail: Victoria.sharpe@tufts.edu

Word count: 3,993

Keywords: lexical predictability; schizophrenia; large language models; computational psychiatry; natural language processing; context; language production

Abstract

Background: Early psychopathologists proposed that certain features of positive thought disorder, the disorganized language output produced by some people with schizophrenia, suggest an insensitivity to global, relative to local, discourse context. This idea has received support from carefully controlled psycholinguistic studies in language comprehension. In language production, researchers have so far remained reliant on subjective qualitative rating scales to assess and understand speech disorganization. Now, however, recent advances in large language models mean that it is possible to quantify sensitivity to global and local context objectively by probing lexical probability (the predictability of a word given its preceding context) during natural language production.

Methods: For each word in speech produced by 60 first-episode psychosis patients and 35 healthy, demographically-matched controls, we extracted lexical probabilities from GPT-3 based on contexts that ranged from very local— a single preceding word: P(Wn | Wn-1)—to global— up to 50 preceding words: P(Wn|Wn-50, Wn-49, …, Wn-1).

Results: We show, for the first time, that disorganized speech is characterized by disproportionate insensitivity to global, versus local, linguistic context. Critically, this global-versus-local insensitivity selectively predicted clinical ratings of positive thought disorder, above and beyond overall symptom severity. There was no evidence of a relationship with negative thought disorder (impoverishment).

Conclusions: We provide an automated, interpretable measure that can potentially be used to quantify speech disorganization in schizophrenia. Our findings directly link the clinical phenomenology of thought disorder to neurocognitive constructs that are grounded in psycholinguistic theory and neurobiology.

Introduction

Since the days of Kraepelin and Bleuler, psychopathologists have struggled to describe and understand the disorganized, incoherent language output produced by some people with schizophrenia — positive thought disorder (1,2). Positive thought disorder affects up to 50% of patients with schizophrenia (3) and is linked to significant impairments in social functioning (4,5) and overall quality of life (3,6). It is therefore of critical clinical importance to understand how to objectively quantify positive thought disorder, and to link this characterization to its underlying mechanisms.

One clue comes from careful descriptions of positive thought disorder; phenomena like tangentiality and derailment may indicate a relative insensitivity to global discourse context. Importantly, this insensitivity seems to be *specific* to global information, with sensitivity to local context being largely intact. In line with this idea, studies across multiple domains have documented reduced sensitivity to global, relative to local, context, in schizophrenia (7-10). Indeed, in language *comprehension*, several tightly controlled psycholinguistic studies, using both behavioral and neural measures, suggest that patients are less able than healthy controls to use global linguistic context (long sentences or discourse) to facilitate the processing of incoming words (11-15), whereas the automatic use of local linguistic context (e.g. directly related semantic primes, short sentence frames) is generally spared (13,16-20). In particular, Swaab and colleagues (13) showed that whereas healthy controls produced a larger N400 (an event-related potential that indexes the probability of words based on their context) in response to discourse-incongruent *versus* discourse-congruent words, people with schizophrenia only showed a congruency effect when the target word was predictable based on its local context.

Until now, however, it has not been possible to objectively quantify the use of global versus local context in natural speech *production* in schizophrenia. In the 1960s and 70s, some researchers attempted to assess patients' general sensitivity to context using fill-in-the-blank "Cloze" completion tasks (e.g., (21-23)), but this procedure was extremely time-consuming and thus impossible to carry out on a large scale. It was also ill-suited for understanding patients' *relative* sensitivity to global versus local context. Therefore, over 110 years after Bleuler's seminal descriptions of disorganized speech in schizophrenia (1), clinicians and researchers remain reliant on subjective qualitative rating scales to assess and understand speech disorganization (positive thought disorder) in schizophrenia (24-30).

Fortunately, due to recent advances in Large Language Models (LLMs), it is now possible to obtain an automated, precise measure that objectively quantifies the relationship between each word and its full (local and global) preceding context — its lexical probability; that is the probability of observing that word, given the full set of words that precedes it: $P(w_n \mid w_1, w_2, \ldots, w_{n-1})$, where $w_1, w_2, \ldots, w_n$ is a sequence of words. *Lexical probability* is the most robust predictor of behavior (31,32) and neural activity (33-38) during language comprehension (for review, see (39)), and it is tightly linked to the coherence of language output (40,41 16410,42-44). Indeed, although not typically designed with biological plausibility in mind, many large language models, like OpenAI's GPT series, are explicitly trained to predict upcoming words based on their preceding context, and in doing so, they learn to produce coherent language that is remarkably human-like (45,46). Moreover, with certain manipulations to GPT's use of context, it produces disorganized, incoherent speech that is very similar to that produced by people with positive thought disorder (47).

Over the past 15 years, various Natural Language Processing (NLP) measures have been

used to achieve high classification accuracy when discriminating between schizophrenia patients and healthy controls--e.g., (48-54); see also (55) for an excellent overview--and predicting psychosis onset in clinical high-risk populations--e.g., (56,52,57 16061)).

While a few of these measures have been shown to correlate with clinical ratings of atypical speech--e.g., (48,58,59)--the degree to which these measures are linked *selectively* to *positive* thought disorder (i.e. language disorganization), as opposed to negative thought disorder (reduced overall production or impoverishment), has been largely unclear (though see (58) for work in clinical high-risk individuals). It is also unclear whether these measures account for variance in thought disorder above and beyond overall symptom severity, or medication use. Therefore, what is badly needed is a selective measure of positive thought disorder that is explicitly grounded in neurocognitive theory and that can not only help us *describe* positive thought disorder, but also *understand* its underlying cognitive and computational mechanisms.

In the present study, we build on this large NLP literature to develop such a measure. We used GPT-3, a state-of-the-art LLM, to estimate the lexical probability of every word in speech produced by untreated first-episode psychosis patients and demographically-matched healthy controls. Critically, we manipulated the length of the context window that the language model had access to, allowing us to obtain graded estimates of predictability for each word based on very local context—a single preceding word: $P(W_n \mid W_{n-1})$—to very global contexts—up to 50 preceding words: $P(W_n|W_{n-50}, W_{n-49}, \ldots, W_{n-1})$. To anticipate our findings, we show for the first time that a disproportionate insensitivity to global versus local context specifically and selectively predicts clinical ratings of positive thought disorder. Thus, we provide an automated, theoretically-grounded measure that links clinical characterizations of thought disorder with a

mechanistic understanding of language production.

Methods

Participants

One-hundred-and-six English-speaking participants (36 healthy controls: HCs; 70 first episode psychosis patients: FEPs) were included from an ongoing study (ClinicalTrials.gov Identifier: NCT02882204). All participants were between 16 and 45 years. Exclusion criteria were: a history of drug or alcohol dependence over the past year, a history of major head injury (with a period of unconsciousness or seizures), intellectual disability, uncontrolled medical conditions, >2 weeks of lifetime antipsychotic exposure, or the inability to provide informed consent.

FEPs were recruited April 2017 – September 2019 by screening all consecutive referrals to the Prevention and Early Intervention for Psychosis Program at the London Health Sciences Centre in London, Ontario, Canada. Patients were approached within two weeks of referral, ensuring that all were in the acute, untreated[1] phase of psychosis. A later six-month consensus diagnosis from two research psychiatrists and the primary treatment provider, based on (61), and the Structured Clinical Interview for DSM-5 (62) indicated that 65 of these participants met criteria for a schizophrenia spectrum disorder and 5 for affective psychosis. The Research Ethics Board at Western University approved all study procedures and all participants provided

[1] The overall study sample had a mean <0.5 defined daily dose equivalents of antipsychotics when speech was assessed (see also (60) for full sample details).

informed consent.

Ten patients and one control participant were subsequently excluded from analysis because of uncertain Parental Socioeconomic Status (PSES (63)) information, which served as an important control variable (see below)[2]. Summary data of the 35 HCs and 60 FEPs included in the reported analyses are shown in Table 1.

<Table 1>

Clinical and neuropsychological assessment

Patients' symptoms were assessed using the 8-Item Positive and Negative Syndrome Scale (PANSS-8 (64); condensed from the PANSS (65) by one of two research psychiatrists, on the same week as speech acquisition (intraclass correlation for total scores (ICC) between the 2 raters; 10 subjects = 0.91).

In all participants, general cognitive function was assessed using three cognitive tasks: (1) the Digit-Symbol Substitution Test (DSST (66)) as a measure of working memory and processing speed; (2) the Category Fluency Task (67) as a measure of semantic memory and executive function, and (3) Part B of the Trail-Making Test (TMT (68)) as a measure of non-verbal executive function. See Supplementary Materials for details.

[2] We note that the pattern of results was the same (a) when we excluded the five patients with affective psychosis, and (b) when we included the ten patients with missing demographic data.

Speech data

All participants described three pictures from the Thematic Apperception Test (69), for one minute each (see Supplementary Materials for details). Their speech was recorded and transcribed, and used for both clinical assessment of thought disorder and extraction of GPT lexical probability measures.

Thought disorder ratings

Thought disorder ratings were completed by a single trained graduate-level research assistant under the supervision of a research psychiatrist, blinded to patient status by using numbered transcripts, using the Thought and Language Index (TLI (30)). For each participant, the rater computed a measure of positive thought disorder ("Disorganization" score) and a measure of negative thought disorder ("Impoverishment" score); see Supplementary Materials for details.

GPT-derived measures of lexical probability

OpenAI's GPT-3 is a LLM with billions of parameters that can provide accurate estimates of the conditional probability of any given word (45), given a sequence of *k* preceding words (the preceding context), i.e. $P(w_n \mid w_{n-k}, w_{n-k+1}, \ldots, w_{n-1})$. We extracted the lexical probability of each word in each participant's speech, based on all available context, as well as based on different

context conditions (see Results; Figure 1). These extracted values served as the dependent measure in a series of linear mixed effects regressions designed to test our a priori hypothesis. To extract these various estimates of lexical probability, we used the 'davinci_002' model via the OpenAI API (https://openai.com/product), after splitting each participant's transcript into contiguous discourse "segments." Within each segment, spellings were standardized and all punctuation except for apostrophes, hyphens, and sentence-final punctuation was removed.

<Figure 1>

Statistical analysis

Prior to statistical analysis, we excluded any probability values for disfluencies (e.g., "um", "uh") and function words. This left a total of 19,421 word tokens (47.72% of the data; Patients: 11,696 tokens, 47.55%; Controls: 7725 tokens, 47.98%). We log-transformed all probability values to ensure that the assumptions of the general linear model were met and to allow us to probe proportional, rather than absolute, differences in lexical probability. We then excluded log probability values outside three standard deviations from the mean of each context condition, which resulted in the exclusion of 81.10 (SD = 15.36) datapoints (i.e. probability values) per condition (see above) on average.

Linear mixed effects regression provides a particularly advantageous analytic approach for this dataset, as it allowed us to test the effects of predictors of interest on Lexical Probability, while accounting for clustering in the data by incorporating by-subject and by-item (word token) random effects. It also allowed us to control for potential confounds by including "nuisance" item-level and participant-level covariates in each model: Segment Length (number of words in

each segment), based on the possibility that participants distribute information differently across discourse segments of different lengths (42-44), and Parental SES (PSES; participant-level), which is associated with variance in various aspects of language use (see (70) for review). All continuous predictor variables were z-scored (see Supplementary Materials for details).

Results

1. People with schizophrenia are less sensitive to context during language production.

We began by asking whether FEPs were generally less sensitive to context than controls. We used GPT-3 to extract the lexical probability of each word in each discourse segment, based on the full set of words the participant had produced up until that point (the prior context)—P(Word | AllAvailableContext). Then, as a baseline, we extracted the lexical predictability of the same words, but replacing the prior context with randomly sampled words from an unrelated picture description (see Supplementary Materials)—P(Word | NoContext).

These lexical probabilities served as the outcome variable in a linear mixed effects model, with predictors of interest ContextType (AllAvailableContext vs. NoContext; within items), Group (FEPs vs. HCs; between participants), and the ContextType*Group interaction; see Table 2, Figure 2. As expected, across groups, participants produced words that were more predictable in the AllAvailableContext condition than in the NoContext condition (a main effect of ContextType). However, this effect was smaller in the patient group (an interaction between Context and Group), with follow-ups showing that, relative to HCs, FEPs produced less predictable words in the AllAvailableContext condition (Est.= -0.38, SE=0.10, $p<.001$), but not the NoContext condition (Est.=-0.07, SE =0.09, $p=0.43$).

<Table 2>

<Figure 2>

2. People with schizophrenia exhibit a *selective* insensitivity to global (versus local) context.

We next turned to the critical question of whether patients are selectively insensitive to global, relative to local, linguistic context.

For each word in each discourse segment, we extracted its lexical probability based on 50 different context window sizes (ranging from 1 preceding word to 50 preceding words; Figure 1), generating up to 50 unique datapoints for each word produced by each participant: 747,195 datapoints total. These values served as the outcome variable in an analysis where the predictors of interest were ContextWindowSize (continuous, within-items) and Group (FEPs vs. HCs; between-participants), and their interaction. ContextWindowSize was log-transformed to produce a more linear relationship with lexical probability (also log-transformed; see Methods), per the assumptions of the general linear model.

As expected, across both groups, larger ContextWindowSize predicted greater lexical probability (a main effect of ContextWindowSize). There was also a main effect of Group: at the mean ContextWindowSize (~22.07 words of context), lexical probability was greater in HCs than in FEPs, consistent with the analysis above. Critically, however, as ContextWindowSize increased (i.e. as the length of the context available to the model increased), lexical probability increased *less* in patients, relative to controls (an interaction between ContextWindowSize and Group); Table 3, Figure 3.

Follow-ups at the local and global extremes of ContextWindowSize confirmed that for more global contexts (averaging across window sizes between 46 to 50 words), the group difference was significant (Est. = -0.25, SE = 0.09, p = 0.01), but for very local contexts (averaging across window sizes between 1-5 words), it was non-significant (Est. = -0.12, SE =

0.08, p = 0.13).

<Table 3>

<Figure 3>

3. Relative insensitivity to global context is not driven by impairments in general cognitive function.

To determine whether the interaction between Group and ContextWindowSize could be explained by differences in overall cognitive functioning, we averaged three scaled scores from each participant's cognitive assessments (see Methods), and included this summary measure (CognitiveFunction), and its interaction with ContextWindowSize, as additional predictors in the above model.

The Group*ContextWindowSize interaction persisted, suggesting that selective insensitivity to global versus local context could not be explained by differences in CognitiveFunction between groups. Indeed, CognitiveFunction failed to predict lexical probability at all (no main effect of CognitiveFunction; no Group*CognitiveFunction interaction; Table 4)[3].

[3] Additional analyses exploring the effects of each cognitive measure separately produced the same results: the interaction between Group and Window Size persisted, whereas there was no main effect, or interaction with Window Size, for any cognitive measure.

<Table 4>

4. Relative insensitivity to global context is selectively associated with positive thought disorder.

We then asked whether, within the patient group, patients' insensitivity to global context was linked to positive thought disorder. To address this question, we carried out an analysis, within the patient group only, with predictors of interest ContextWindowSize (continuous, within-items), Disorganization (TLI subscore; continuous, between-participants), and their interaction. This revealed a significant interaction between ContextWindowSize and Disorganization, such that Disorganization predicted the effect of ContextWindowSize on lexical probability. Specifically, greater Disorganization was associated with a smaller increase in lexical probability as ContextWindowSize increased; Table 5, Figure 4.

Follow-ups at the local and global extremes of ContextWindowSize confirmed that for very local contexts (averaging across window sizes between 1-5 words), there was no effect of Disorganization (Est.=0.01, SE=0.04, p=.64), but for more global contexts (averaging across window sizes between 46 to 50 words), Disorganization significantly predicted lexical probability (Est.=-.11, SE=0.04, p=0.02).

<Table 5>

<Figure 4>

To determine whether global-vs-local insensitivity was specifically linked to *positive* thought disorder, as opposed to negative thought disorder, we then ran an equivalent analysis in which we replaced Disorganization with the Impoverishment, as a measure of negative thought disorder. This revealed only a trending main effect of Impoverishment, and no Impoverishment*ContextWindowSize interaction; Table 6.

<Table 6>

Finally, to determine whether there was an effect of Disorganization over and above overall symptom severity, we ran an additional analysis which also included the PANSS-8 Total score and its interaction with ContextWindowSize (Table 7). The significant interaction between Disorganization and ContextWindowSize persisted. In contrast, neither the main effect of PANSS-8 nor the PANSS-8*ContextWindowSize interaction was significant.

<Table 7>

Discussion

We used GPT-3, a state-of-the art large language model, to show that the speech produced by a large group of first-episode psychosis patients is selectively insensitive to global, relative to local, context. This effect of global-vs.-local insensitivity specifically predicted severity of positive thought disorder (disorganized speech). In contrast, we saw no evidence of a relationship with negative thought disorder (impoverishment) and no relationship with overall symptom severity. These findings have important clinical and theoretical implications.

Clinical implications: A measure for quantifying positive thought disorder

From a clinical perspective, our index of relative sensitivity to global vs. local context provides an objective, automatic, interpretable measure that can potentially be used to quantify speech disorganization in schizophrenia. The relationship between global-vs-local insensitivity and positive thought disorder was selective. It was also graded: at very local context windows (up to about five words), there was no relationship between lexical probability and positive thought disorder. However, as the window size increased, differences in lexical probability became increasingly large with greater severity of positive thought disorder. In future work, it will be important to extend these findings to patients in the chronic phase of illness and to precisely test the psychometric properties of this measure. With further testing and development, we suggest that it could assist in various clinical purposes, ranging from diagnosis to symptom monitoring. It might also be able to detect subtle language production atypicalities that are less clinically obvious, but that nonetheless impact real-world communicative function. Finally, although the present study focused exclusively on schizophrenia, the methods we have used are

very amenable to a transdiagnostic approach, thus laying groundwork for examination of language impairments across diagnostic categories (e.g. bipolar disorder).

Theoretical implications for a mechanistic understanding of positive thought disorder

Crucially, this measure of global-vs.-local insensitivity goes beyond description. It directly links the clinical phenomenology of thought disorder to neurocognitive constructs that are grounded in psycholinguistic theory.

In healthy adults, numerous psycholinguistic studies of language comprehension have established that healthy adults continually track and use all available contextual information to facilitate lexical processing in real time, as evidenced by both neural (33-38) and behavioral (31,32) measures. In schizophrenia, there is evidence that, although patients' ability to establish and use local dependencies (world-level priming, clauses, or short sentences) is largely intact (13,16-20), their use of more global sources of information (longer sentences, discourse, or even high-level visual context) during language comprehension is impaired (12-14,16,15). Here, we show, for the first time, that this same global-vs-local insensitivity characterizes natural language *production* in schizophrenia and specifically predicts clinical ratings of positive thought disorder, raising the possibility that atypicalities in language comprehension and production in schizophrenia are driven by shared neurocognitive mechanisms.

These atypicalities are not reducible to the type of generalized cognitive deficits that can impede patients' performance on challenging neuropsychological tasks (see (71,72)). In healthy adults, the ability to use global context to inform word-by-word language processing is not effortful; rather, it occurs implicitly, without conscious effort. Indeed, in the present study, we

found no evidence that patients' relative insensitivity to global context could be explained by a deficit in overall cognitive function, as indexed by their performance in the neuropsychological tasks we administered. On the other hand, it will be important for future studies to examine the relationship between the atypicalities described here and patients' performance in tasks that have been linked to selective insensitivities to global context in other cognitive and perceptual domains (e.g., (73-75)) as well as to impairments in the use of context more generally (e.g., (76,77)).

From a computational perspective, selective insensitivity to global linguistic context can be understood within a hierarchical generative framework: a general theory, based on the principles of Bayesian updating, of how brain perceives, interprets and acts upon the world (78,79). Aberrations in hierarchical generative circuits have been proposed as a holistic explanation for multiple neurocognitive atypicalities and symptoms in schizophrenia (80-83), including in language (84,85).

Such a framework naturally explains why the healthy brain is so sensitive to a word's probability given its prior local and global context (39,86). That is, effective communication requires both the producer and comprehender to employ and continually update hierarchically-organized internal generative models that represent information over successively longer time scales. Individual words (lexical representations) are encoded at relatively short time scales at lower levels of the hierarchy; local semantic/syntactic dependencies are encoded at medium timescales at middle levels; and broader semantic structures (e.g., whole topics and situational contexts) are encoded at the longest timescales at the highest levels of the hierarchy. Thus, within this system, effects of context on lexical probability emerge as a byproduct of optimal

communication: as each word is produced/processed in real time, lexical representations serve as a "causal bottleneck", informed by statistical information encoded at all levels above (31).

In contrast to the human brain, large language models like GPT do not implement probabilistic inference. Rather, they are explicitly trained to predict upcoming words based on vast quantities of human text. Here, we leveraged this property to obtain estimates of lexical probability for each word in patients' utterances, given local and global context, demonstrating that the disorganized speech produced by first-episode schizophrenia patients fails to benefit from the additional weighting and constraints typically conferred by global information. These findings pinpoint patients' deficits in linguistic context processing during language production to the highest levels of the generative hierarchy (see (84) for discussion). However, they also raise many additional questions and open up important avenues for future research.

First, what is the precise nature of this high-level atypicality? One possibility is that patients cannot maintain stable representations at the top of their generative hierarchy; that is, at any given time, they may be uncertain about the underlying topic. This type of high-level "belief instability" (87) intuitively explains phenomena that characterize disorganized speech like tangentiality and derailment. Another possibility, however, is that patients fail to flexibly *update* high-level representations (88) (although overly rigid high-level beliefs might also be expected to result in an insensitivity to local context).

GPT is not well-suited for distinguishing these possibilities. Because of its black box nature, we know little about the internal representations it learns. An alternative approach would be to employ language models that are specifically trained to capture latent high-level topic representations from speech outputs (e.g., BERTopic (89)). These topic models could be used to

quantify topic diversity within speech samples, allowing us to determine whether uncertainty over topic representations is excessively high in patients' speech compared to that of healthy controls.

A second set of open questions concerns the specific neurocomputational mechanisms by which atypicalities at the highest levels of representation give rise to selective lexical-level insensitivity to global context during real-time language production and comprehension. Again, most current large language models are ill-suited for addressing this question. Whereas their architectures are feedforward in nature, the human brain is characterized by long-range feedback connections that bridge the highest and lowest levels of the cortical hierarchy. These connections are deeply integrated within the cortical microcircuitry (90), allowing for a continuous interactive exchange of information during word-by-word processing (91).

Understanding how these neural dynamics are affected in schizophrenia would therefore require yet another type of computational model, whose architecture is constrained by what we know about neurobiology. One such model is *predictive coding* — a specific implementation of the more general hierarchical generative framework described above (92-94). Predictive coding is biologically plausible (95-97) and has been able to simulate multiple neural phenomena in human perception, both in healthy adults (92,93,98), as well as in schizophrenia (99).

Indeed, in recent work, we have shown that, in healthy individuals, predictive coding is able to simulate probabilistic effects of context (i.e., effects of lexical probability) on neural activity (100). It can also explain the time course and localization of neural activity produced by incoming words across the left-lateralized fronto-temporal hierarchy during language processing (101). Therefore, an important goal of future studies will be to determine whether the dynamics

of predictive coding can explain the specific abnormalities in fronto-temporal neural activity that are commonly observed in thought disorder (for review, see (102)).

Conclusion

We show that patients in the acute phase of schizophrenia are selectively insensitive to global contextual information during language production. These findings link an objective, theory-driven measure of contextual sensitivity to clinical assessments of positive thought disorder, laying groundwork for an understanding of the computational mechanisms and neural circuitry underlying disorganized language production. Thus, we have begun to realize Bleuler's original proposal that the basic phenomenology of thought disorder can be linked to the fundamental mechanisms of schizophrenia.

Acknowledgments

This study was funded by the Sidney Baer Trust. Data acquisition was funded by the Canadian Institutes of Health Research (CIHR) Foundation Grant (Grant no. 375104/2017) to Lena Palaniyappan. Lena Palaniyappan acknowledges research support from the Canada First Research Excellence Fund, awarded to the Healthy Brains, Healthy Lives initiative at McGill University (through New Investigator Supplement to Lena Palaniyappan); Monique H. Bourgeois Chair in Developmental Disorders and Graham Boeckh Foundation (Douglas Research Centre, McGill University) and salary award from the Fonds de recherche du Quebec-Santé (FRQS). We wish to thank Hannah Ke, Sabrina Ford, Cassandra Branco (for data curation and transcription support), Drs. Ross Norman, Peter Liddle (for training support), Drs. Julie Richard, Hooman Ganjavi, Priya Subramanian and Raj Harricharan (for clinical referrals).We are also grateful to undergraduate research assistants Emma Draisin, Brooke Brody, Zoe Gardaret, and Maggie Wargo – and particularly Santiago Noriega, Feng Cheng, and Julia Klein - for their help with data processing and analysis.

Disclosures

The authors have no conflicts of interest to report.

References

1. Bleuler, E. (1911/1950). *Dementia Praecox, or the Group of Schizophrenias* (J. Zinker, Trans.). New York: International Universities Press.
2. Kraepelin, E. (1971). *Dementia Praecox and Paraphrenia* (E. Barclay, Trans.). New York: Krieger.
3. Breier, A., & Berg, P. H. (2003). The psychosis of schizophrenia. *Biological Psychiatry, 46*, 361-364.
4. Marggraf, M. P., Lysaker, P. H., Salyers, M. P., & Minor, K. S. (2020). The link between formal thought disorder and social functioning in schizophrenia: A meta-analysis. *European Psychiatry, 63*(1), e34. doi:10.1192/j.eurpsy.2020.30
5. Bowie, C. R., & Harvey, P. D. (2008). Communication abnormalities predict functional outcomes in chronic schizophrenia: Differential associations with social and adaptive functions. *Schizophrenia Research, 103*(1-3), 240-247.
6. Tan, E. J., Thomas, N., & Rossell, S. L. (2014). Speech disturbances and quality of life in schizophrenia: differential impacts on functioning and life satisfaction. *Comprehensive Psychiatry, 55*(3), 693-698. doi:10.1016/j.comppsych.2013.10.016
7. Silverstein, S. M., Knight, R. A., Schwarzkopf, S. B., West, L. L., Osborn, L. M., & Kaminen, D. (1996). Stimulus configuration and context effects in perceptual organization in schizophrenia. *Journal of Abnormal Psychology, 105*(3), 410-420. doi:10.1037/0021-843x.105.3.410
8. Hemsley, D. R. (2005). The development of a cognitive model of schizophrenia: Placing it in context. *Neuroscience and Biobehavioral Reviews, 29*(6), 977-988. doi:10.1016/j.neubiorev.2004.12.008

9. Bellgrove, M. A., Vance, A., & Bradshaw, J. L. (2003). Local-global processing in early-onset schizophrenia: evidence for an impairment in shifting the spatial scale of attention. *Brain and Cognition, 51*(1), 48-65. doi:10.1016/s0278-2626(02)00509-2
10. Silverstein, S. M. (2016). Visual Perception Disturbances in Schizophrenia: A Unified Model. In M. Li & W. D. Spaulding (Eds.), *The Neuropsychopathology of Schizophrenia: Molecules, Brain Systems, Motivation, and Cognition* (pp. 77-132). Cham: Springer International Publishing.
11. Sitnikova, T., Salisbury, D. F., Kuperberg, G. R., & Holcomb, P. J. (2002). Electrophysiological insights into language processing in schizophrenia. *Psychophysiology, 39*(6), 851-860.
12. Ditman, T., & Kuperberg, G. R. (2007). The time course of building discourse coherence in schizophrenia: An ERP investigation. *Psychophysiology, 44*(6), 991-1001. doi:10.1111/j.1469-8986.2007.00565.x
13. Swaab, T. Y., Boudewyn, M. A., Long, D. L., Luck, S. J., Kring, A. M., Ragland, J. D., . . . Carter, C. S. (2013). Spared and impaired spoken discourse processing in schizophrenia: effects of local and global language context. *Journal of Neuroscience, 33*(39), 15578-15587. doi:10.1523/jneurosci.0965-13.2013
14. Boudewyn, M. A., Carter, C. S., Long, D. L., Traxler, M. J., Lesh, T. A., Mangun, G. R., & Swaab, T. Y. (2017). Language context processing deficits in schizophrenia: The role of attentional engagement. *Neuropsychologia, 96*, 262-273.
15. Titone, D., Levy, D. L., & Holzman, P. S. (2000). Contextual insensitivity in schizophrenic language processing: Evidence from lexical ambiguity. *Journal of*

Abnormal Psychology, 109(4), 761-767. query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11196002

16. Rabagliati, H., Delaney-Busch, N., Snedeker, J., & Kuperberg, G. R. (2018). Spared bottom-up but impaired top-down interactive effects during naturalistic language processing in schizophrenia: Evidence from the visual world paradigm. *Psychological Medicine*, 1-11. doi:10.1017/S0033291718001952
17. Kuperberg, G. R. (2010). Language in schizophrenia Part 1: an Introduction. *Language and Linguistics Compass, 4*(8), 576-589. doi:10.1111/j.1749-818X.2010.00216.x
18. Minzenberg, M. J., Ober, B. A., & Vinogradov, S. (2002). Semantic priming in schizophrenia: A review and synthesis. *Journal of the International Neuropsychological Society, 8*(5), 699-720.
19. Andrews, S., Shelley, A., Ward, P. B., Fox, A., Catts, S. V., & McConaghy, N. (1993). Event-related potential indices of semantic processing in schizophrenia. *Biological Psychiatry, 34*, 443-458
20. Kuperberg, G. R., Sitnikova, T., Goff, D., & Holcomb, P. J. (2006). Making sense of sentences in schizophrenia: electrophysiological evidence for abnormal interactions between semantic and syntactic processing. *Journal of Abnormal Psychology, 115*(2), 251-265. doi:10.1037/0021-843X.115.2.251
21. Manschreck, T. C., Maher, B. A., Rucklos, M. E., & White, M. T. (1979). The predictability of thought disordered speech in schizophrenic patients. *British Journal of Psychiatry, 134*, 595-601.

22. Ragin, A. B., & Oltmanns, T. F. (1983). Predictability as an index of impaired verbal communication in schizophrenic and affective disorders. *The British Journal of Psychiatry, 143*(6), 578-583.

23. Ragin, A. B., & Oltmanns, T. F. (1987). Communicability and thought disorder in schizophrenics and other diagnostic groups: A follow-up study. *The British Journal of Psychiatry, 150*(4), 494-500.

24. Andreasen, N. C. (1979). Thought, language and communication disorders. I. Clinical assessment, definition of terms, and evaluation of their reliability. *Archives of General Psychiatry, 36*, 1315-1321.

25. Andreasen, N. C. (1986). Scale for assessment of thought, language and communication (TLC). *Schizophrenia Bulletin, 12*, 473-482.

26. Johnston, M. H., & Holtzman, P. S. (1979). *Assessing Schizophrenic Thinking*. San Francisco: Jossey-Bass.

27. Solovay, M. R., Shenton, M. E., Gasperetti, C., Coleman, C., Kestenbaum, J., Carpenter, T., & Holtzman, P. S. (1986). Scoring manual for the thought disorder index. *Schizophrenia Bulletin, 12*, 483-496.

28. Marengo, J. T., Harrow, M., Lanin-Kettering, I., & Wilson, A. (1985). A manual for assessing aspects of bizarre-idiosyncratic thinking. In M. Harrow & D. Quinlan (Eds.), *Disordered Thinking and Schizophrenic Psychopathology* (pp. 394-411). New York: Gardner Press.

29. Marengo, J. T., Harrow, M., Lanin-Kettering, I., & Wilson, A. (1986). Evaluating bizzarre-idiosyncratic thinking: A comprehensive index of positive thought disorder. *Schizophrenia Bulletin, 12*, 497-509.

30. Liddle, P. F., Ngan, E., Caissie, S. L., Anderson, C. M., Bates, A. T., Quested, D. J., . . . Weg, R. (2002). Thought and Language Index: an instrument for assessing thought and language in schizophrenia. *British Journal of Psychiatry, 181*(4), 326-330. doi:10.1192/bjp.181.4.326

31. Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*(3), 302-319. doi:10.1016/j.cognition.2013.02.013

32. Brothers, T., & Kuperberg, G. R. (2021). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. *Journal of Memory and Language, 116*. doi:10.1016/j.jml.2020.104174

33. Michaelov, J. A., Coulson, S., & Bergen, B. K. (2021). So Cloze yet so Far: N400 amplitude is better predicted by distributional information than human predictability judgements. *arXiv preprint arXiv:2109.01226*.

34. Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language, 123*. doi:10.1016/j.jml.2021.104311

35. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 119*(32), e2201968119. doi:10.1073/pnas.2201968119

36. Caucheteux, C., Gramfort, A., & King, J. R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*. doi:10.1038/s41562-022-01516-2

37. Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. *Preparatory States and Processes*, 217-237.

38. DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121. doi:10.1038/nn1504

39. Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32-59. doi:10.1080/23273798.2015.1102299

40. Taylor, W. (1953). 'Cloze' procedure: A new tool for measuring readability. *Journalism Quarterly, 30*, 415-433.

41. Oller, J. W., & Jonz, J. (1994). *Cloze and Coherence*: Associated University Presses.

42. Genzel, D., & Charniak, E. (2002). *Entropy rate constancy in text.* Paper presented at the Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA.

43. Genzel, D., & Charniak, E. (2003). *Variation of entropy and parse trees of sentences as a function of the sentence number.* Paper presented at the Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Sapporo, Japan.

44. Meister, C., Pimentel, T., Haller, P., Jager, L., Cotterell, R., & Levy, R. (2021). Revisiting the Uniform Information Density Hypothesis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*.

45. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 1877-1901. doi:10.5555/3495724.3495883

46. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog, 1*(8), 9.

47. Fradkin, I., Nour, M. M., & Dolan, R. J. (2023). Theory-Driven Analysis of Natural Language Processing Measures of Thought Disorder Using Generative Language Modeling. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 8*(10), 1013-1023. doi:10.1016/j.bpsc.2023.05.005

48. Elvevåg, B., Foltz, P. W., Weinberger, D. R., & Goldberg, T. E. (2007). Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophrenia Research, 93*(1-3), 304-316.

49. Mota, N. B., Vasconcelos, N. A., Lemos, N., Pieretti, A. C., Kinouchi, O., Cecchi, G. A., . . . Ribeiro, S. (2012). Speech graphs provide a quantitative measure of thought disorder in psychosis. *PloS One, 7*(4), e34928. doi:10.1371/journal.pone.0034928

50. Mota, N. B., Copelli, M., & Ribeiro, S. (2017). Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *NPJ Schizophrenia, 3*, 18. doi:10.1038/s41537-017-0019-3

51. Mitchell, M., Hollingshead, K., & Coppersmith, G. (2015). *Quantifying the language of schizophrenia in social media.* Paper presented at the Proceedings of the 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality.

52. Corcoran, C. M., Carrillo, F., Fernandez-Slezak, D., Bedi, G., Klim, C., D. C. Jacitt, . . . Cecchi, G. A. (2018). Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry, 17*(1), 67-75.

53. Iter, D., Yoon, J., & Jurafsky, D. (2018). *Automatic detection of incoherent speech for diagnozing schizophrenia*. Paper presented at the Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, New Orleans, LA.

54. Voleti, R., Woolridge, S., Liss, J. M., Milanovic, M., Bowie, C. R., & Berisha, V. (2019). Objective assessment of social skills using automated language analysis for identification of schizophrenia and bipolar disorder. *arXiv preprint, arXiv: 1904.10622*.

55. Corcoran, C. M., Mittal, V. A., Bearden, C. E., R, E. G., Hitczenko, K., Bilgrami, Z., . . . Wolff, P. (2020). Language as a biomarker for psychosis: A natural language processing approach. *Schizophrenia Research, 226*, 158-166. doi:10.1016/j.schres.2020.04.032

56. Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., . . . Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophrenia, 1*, 15030.

57. Spencer, T. J., Thompson, B., Oliver, D., Diederen, K., Demjaha, A., Weinstein, S., . . . McGuire, P. (2021). Lower speech connectedness linked to incidence of psychosis in people at clinical high risk. *Schizophrenia Research, 228*, 493-501. doi:10.1016/j.schres.2020.09.002

58. Bilgrami, Z. R., Sarac, C., Srivastava, A., Herrera, S. N., Azis, M., Haas, S. S., . . . Corcoran, C. M. (2022). Construct validity for computational linguistic metrics in

individuals at clinical risk for psychosis: Associations with clinical ratings. *Schizophr Res, 245*, 90-96. doi:10.1016/j.schres.2022.01.019

59. Tang, S. X., Hansel, K., Cong, Y., Nikzad, A. H., Mehta, A., Cho, S., . . . Liberman, M. Y. (2023). Latent Factors of Language Disturbance and Relationships to Quantitative Speech Features. *Schizophrenia Bulletin, 49*(Suppl_2), S93-S103. doi:10.1093/schbul/sbac145
60. Dalal, T. C., Liang, L., Silva, A. M., Mackinley, M., Voppel, A., & Palaniyappan, L. (2024). Speech based natural language profile before, during and after the onset of psychosis: A cluster analysis. *Acta Psychiatrica Scandinavica*. doi:10.1111/acps.13685
61. Leckman, J. F., Sholomskas, D., Thompson, D., Belanger, A., & Weissman, M. M. (1982). Best estimate of lifetime psychiatric diagnosis. *Archives of General Psychiatry, 39*(8), 879-883.
62. First, M. B. (2014). Structured clinical interview for the DSM (SCID). In *The encyclopedia of clinical psychology* (pp. 1-6).
63. Rose, D., Pevalin, D. J., & O'Reilly, K. *Rose, David, David J. Pevalin, and Karen O'Reilly. The National Statistics Socio-economic Classification: origins, development and use.*: Palgrave Macmillan.
64. Lin, C. H., Lin, H. S., Lin, S. C., Kuo, C. C., Wang, F. C., & Huang, Y. H. (2018). Early improvement in PANSS-30, PANSS-8, and PANSS-6 scores predicts ultimate response and remission during acute treatment of schizophrenia. *Acta Psychiatrica Scandinavica, 137*(2), 98-108. doi:10.1111/acps.12849
65. Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin, 13*(2), 261-276.

66. Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale, Revised*. New York: Psychological Corporation.

67. Neill, E., Gurvich, C., & Rossell, S. L. (2014). Category fluency in schizophrenia research: is it an executive or semantic measure? *Cognitive Neuropsychiatry, 19*(1), 81-95. doi:10.1080/13546805.2013.807233

68. Arbuthnott, K., & Frank, J. (2000). Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology, 22*(4), 518-528. doi:10.1076/1380-3395(200008)22:4;1-0;FT518

69. Murray, H. A. (1943). Thematic apperception test.

70. Palaniyappan, L. (2021). More than a biomarker: could language be a biosocial marker of psychosis? *NPJ Schizophr, 7*(1), 42. doi:10.1038/s41537-021-00172-1

71. Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychol Bull, 79*(6), 380-385. doi:10.1037/h0034541

72. Gold, J. M., & Dickinson, D. (2012). "Generalized cognitive deficit" in schizophrenia: overused or underappreciated? *Schizophrenia Bulletin*, sbs143.

73. Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. *Frontiers in Psychology, 4*, 60. doi:10.3389/fpsyg.2013.00060

74. Boudewyn, M. A., Carter, C. S., & Swaab, T. Y. (2012). Cognitive control and discourse comprehension in schizophrenia. *Schizophrenia Research and Treatment, 2012*, 484-502. doi:10.1155/2012/484502

75. Silverstein, S. M., & Keane, B. P. (2011). Perceptual organization impairment in schizophrenia and associated brain mechanisms: review of research from 2005 to 2010. *Schizophrenia Bulletin, 37*(4), 690-699. doi:10.1093/schbul/sbr052

76. Barch, D. M., & Ceaser, A. (2012). Cognition in schizophrenia: core psychological and neural mechanisms. *Trends in Cognitive Sciences, 16*(1), 27-34. doi:10.1016/J.Tics.2011.11.015

77. Cohen, J. D., Barch, D. M., Carter, C., & Servan-Schreiber, D. (1999). Context-processing deficits in schizophrenia: Converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology, 108*(1), 120-133.

78. Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1456), 815-836. doi:10.1098/Rstb.2005.1622

79. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204. doi:10.1017/S0140525X12000477

80. Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience, 10*(1), 48-58.

81. Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry, 4*, 47. doi:10.3389/fpsyt.2013.00047

82. Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., . . . Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry, 84*(9), 634-643. doi:10.1016/j.biopsych.2018.05.015

83. Frith, C. D., & Friston, K. J. (2013). False perceptions & false beliefs: Understanding schizophrenia. In *Neurosciences and the Human Person: New Perspectives on Human Activities* (Vol. 121, pp. 1-15). Vatican City, Italy: The Pontifical Academy of Sciences.

84. Brown, M., & Kuperberg, G. R. (2015). A hierarchical generative framework of language processing: Linking language perception, interpretation, and production abnormalities in schizophrenia. *Frontiers in Human Neuroscience, 9*, 643. doi:10.3389/fnhum.2015.00643

85. Palaniyappan, L., & Venkatasubramanian, G. (2022). The Bayesian brain and cooperative communication in schizophrenia. *Journal of Psychiatry and Neuroscience, 47*(1), E48-E54. doi:10.1503/jpn.210231

86. Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177. doi:10.1016/j.cognition.2007.05.006

87. Adams, R. A., Napier, G., Roiser, J. P., Mathys, C., & Gilleen, J. (2018). Attractor-like dynamics in belief updating in schizophrenia. *Journal of Neuroscience, 38*(44), 9471-9485. doi:10.1523/JNEUROSCI.3163-17.2018

88. Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science, 357*(6351), 596-600. doi:10.1126/science.aan3458

89. Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv*, arXiv:2203.05794. doi:10.48550/arXiv.2203.05794

90. Douglas, R. J., Martin, K. A. C., & Whitteridge, D. (1989). A Canonical Microcircuit for Neocortex. *Neural Computation, 1*(4), 480-488. doi:10.1162/neco.1989.1.4.480

91. MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676-703. doi:10.1037/0033-295X.101.4.676

92. Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79-87. doi:10.1038/4580

93. Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing, 17*(3), 279-305. doi:10.1007/s10339-016-0765-6

94. Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision Research, 48*(12), 1391-1408. doi:10.1016/j.visres.2008.03.009

95. Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695-711. doi:10.1016/j.neuron.2012.10.038

96. Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology, 7*, 1792. doi:10.3389/fpsyg.2016.01792

97. Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics, 66*(3), 241-251. doi:10.1007/BF00198477

98. Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience, 32*(11), 3665-3678. doi:10.1523/Jneurosci.5003-11.2012

99. Wacongne, C. (2016). A predictive coding account of MMN reduction in schizophrenia. *Biol Psychol, 116*, 68-74. doi:10.1016/j.biopsycho.2015.10.011

100. Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2024). A predictive coding model of the N400. *Cognition, 246*, 105755. doi:10.1016/j.cognition.2024.105755

101. Wang, L., Schoot, L., Brothers, T., Alexander, E., Warnke, L., Kim, M., . . . Kuperberg, G. R. (2023). Predictive coding across the left fronto-temporal hierarchy during language comprehension. *Cerebral Cortex, 33*(8), 4478-4497. doi:10.1093/cercor/bhac356

102. Kircher, T., Bröhl, H., Meier, F., & Engelen, J. (2018). Formal thought disorders: from phenomenology to neurobiology. *The Lancet Psychiatry*. doi:10.1016/s2215-0366(18)30059-2

Figure Captions

Figure 1. Process for extracting GPT lexical probabilities from participant speech samples given varying amounts of context. *(1)* Speech is transcribed, and spellings and punctuation are standardized. *(2)* For a given word *w*, we compute its lexical probability given 50 different context lengths, ranging from one word $P(w_n \mid w_{n-1})$ to 50 words $P(w_n \mid w_{n-50})$. *(3)* We repeat these computations for each word in each contiguous speech segment for each participant. These probabilities can then be used as single datapoints in mixed effects linear regression models, or averaged across context window sizes and groups, as in Figure 3.

Figure 2. Mean lexical predictability, by context type (NoContext vs. AllAvailableContext) for controls vs. patients. Error bars represent standard error.

Figure 3. Mean lexical predictability, by ContextWindowSize (ranging from 1 to 50 words) for controls vs. patients. Shaded areas represent standard error.

Figure 4. Relationship between speech disorganization, as measured by the TLI Disorganization sub-score (larger score indicates greater speech disorganization), and by-subject local/global bias (i.e. the by-subject slopes for the ContextWindowSize on lexical probability) within patients. Black line represents the regression line of best fit.

Tables

Demographics and Clinical Measures

	Controls	Patients	
N	35	60	
Age	21.68 (3.22)	22.26 (4.34)	t(87.56) = -0.74; p = .46
Parental SES[a]	3.09 (1.36)	3.50 (1.27)	t(67.36) = -1.47; p = .15
Sex[b]	F:12; M: 23 ; NB: 0	F: 13; M: 47; NB: 0	X^2 (1) = 1.22; p = .27
PANSS-8 Total[c]	**8.00 (0.00)**	**25.79 (7.08)**	**t(56) = -18.97; p < .05**
PANSS-8 Positive[c]	**3.00 (0.00)**	**12.48 (3.14)**	**t(57) = -22.95; p < .05**
PANSS-8 Negative[c]	**3.00 (0.00)**	**7.48 (4.45)**	**t(57) = -7.68; p < .05**
TLI Total[d]	**0.31 (0.40)**	**1.6 (1.43)**	**t(74.15) = -6.50; p < .05**
TLI Impoverishment[d]	**0.14 (0.25)**	**0.58 (0.72)**	**t(80.62) = -4.20; p < .05**
TLI Disorganization[d]	**0.17 (0.26)**	**1.02 (1.26)**	**t(67.97) = -5.07; p < .05**
DSST Score[e]	**69.20 (11.00)**	**50.20 (13.30)**	**t(82.04) = 7.49; p < .05**
Category Fluency (N Exemplars)[f]	**24.70 (7.01)**	**18.20 (5.12)**	**t(51.99) = 4.41; p < .05**
TMT Part B Score[g]	**55.10 (14.90)**	**96.40 (74.40)**	**t(45.40) = -3.50; p < .05**
Segment Length[h]	**99.41 (35.23)**	**67.47 (31.63)**	**t(65.17) = 4.42; p < .05**

Table 1. By-group means with standard deviations in parentheses; significant differences between groups shown in bold. [a](63); [b]F = Female; M = Male; NB = Non-Binary/Intersex; [c](64,65); [d](30); [e]Digit-Symbol Substitution Test (66), mean of Written and Oral scores; [f]Category/Semantic Fluency: number of animal exemplars produced; [g]Trail-Making Test, Part B (68); [h]Mean length of contiguous speech segment (in words)

Effect of Group and ContextType (NoContext vs. AllAvailableContext) on Lexical Probability

	Estimate	SE	t	p	Sig.
Intercept	-6.54	0.04	-158.12	0.00	***
Group	**-0.23**	**0.08**	**-2.72**	**0.01**	**
Context Type	**4.63**	**0.04**	**109.10**	**0.00**	***
PSES	-0.05	0.04	-1.32	0.19	
SegmentLength	0.34	0.02	17.06	0.00	***
ContextType*Group	**-0.29**	**0.09**	**-3.35**	**0.00**	**
ContextType*PSES	-0.02	0.04	-0.45	0.65	
ContextType*SegmentLength	0.17	0.04	4.75	0.00	***

Table 2. Predictors of interest are shown in bold.

Effect of Group and ContextWindowSize on Lexical Probability

	Estimate	SE	t	p	Sig.
Intercept	-4.65	0.04	-106.33	0.00	***
ContextWindowSize	**0.81**	**0.01**	**82.83**	**0.00**	***
Group	**-0.34**	**0.09**	**-3.75**	**0.00**	***
PSES	-0.07	0.04	-1.74	0.09	.
SegmentLength	-0.03	0.03	-1.32	0.19	
ContextWindowSize*Group	**-0.08**	**0.02**	**-4.04**	**0.00**	***
ContextWindowSize*PSES	-0.01	0.01	-0.61	0.54	
ContextWindowSize*SegmentLength	-0.03	0.01	-4.20	0.00	***

Table 3. Predictors of interest are shown in bold.

Effect of Overall Cognition, Group, and ContextWindowSize on Lexical Probability

	Estimate	SE	t	p	Sig.
Intercept	-4.50	0.08	-58.57	0.00	***
CognitiveFunction	**0.01**	**0.06**	**0.21**	**0.83**	
ContextWindowSize	**0.84**	**0.02**	**48.11**	**0.00**	***
Group	**-0.31**	**0.11**	**-2.78**	**0.01**	**
PSES	-0.09	0.05	-1.73	0.09	.
SegmentLength	-0.04	0.03	-1.21	0.23	
ContextWindowSize*CognitiveFunction	**0.00**	**0.01**	**-0.26**	**0.80**	
ContextWindowSize*Group	**-0.08**	**0.03**	**-3.24**	**0.00**	**
ContextWindowSize*PSES	-0.01	0.01	-0.89	0.38	
ContextWindowSize*SegmentLength	-0.03	0.01	-3.39	0.00	***

Table 4. Predictors of interest are shown in bold.

Effect of TLI Disorganization and ContextWindowSize on Lexical Probability (Within Patients)

	Estimate	SE	t	p	Sig.
Intercept	-4.79	0.06	-79.74	0.00	***
Disorganization	**-0.08**	**0.04**	**-2.00**	**0.05**	**.**
ContextWindowSize	**0.78**	**0.01**	**59.00**	**0.00**	***
Segment_Length	-0.05	0.04	-1.47	0.14	
PSES	-0.10	0.06	-1.82	0.07	.
ContextWindowSize*Disorganization	**-0.02**	**0.01**	**-2.75**	**0.01**	**
ContextWindowSize*Segment_Length	-0.03	0.01	-3.28	0.00	**
ContextWindowSize*PSES	0.01	0.01	0.54	0.59	

Table 5. Predictors of interest are shown in bold.

Effect of TLI Impoverishment and ContextWindowSize on Lexical Probability (Within Patients)

	Estimate	SE	t	p	Sig.
Intercept	-4.81	0.06	-83.12	0.00	***
Impoverishment	**-0.09**	**0.05**	**-1.77**	**0.08**	**.**
ContextWindowSize	**0.77**	**0.01**	**57.84**	**0.00**	***
Segment_Length	-0.07	0.04	-1.89	0.06	**.**
PSES	-0.10	0.06	1.81	0.08	.
ContextWindowSize* Impoverishment	**0.00**	**0.01**	**-0.35**	**0.73**	
ContextWindowSize*Segment_Length	0.03	0.01	-3.54	0.00	***
ContextWindowSize*PSES	0.01	0.01	0.54	0.59	

Table 6. Predictors of interest are shown in bold.

Effect of TLI Disorganization and ContextWindowSize on Lexical Probability Over and Above PANSS-8 (Within Patients)

	Estimate	SE	t	p	Sig.
(Intercept)	-4.86	0.08	-60.85	0.00	***
Disorganization	**-0.07**	**0.04**	**-1.70**	**0.10**	**.**
ContextWindowSize	**0.78**	**0.02**	**44.42**	**0.00**	***
PSES	-0.10	0.06	-1.75	0.09	.
SegmentLength	-0.04	0.04	-1.22	0.22	
PANSS-8	**0.07**	**0.08**	**0.93**	**0.36**	
Disorganization*ContextWindowSize	**-0.02**	**0.01**	**-2.80**	**0.01**	**
ContextWindowSize*PSES	0.01	0.01	0.92	0.36	
ContextWindowSize*SegmentLength	-0.03	0.01	-3.25	0.00	**
ContextWindowSize*PANSS-8	**0.00**	**0.02**	**0.15**	**0.88**	

Table 7. Predictors of interest are shown in bold.

1

Picture Description Sample

*"It looks like the man is sitting on a blanket in the park and it seems like behind him there are dogs running and people talking with leashes in their hands and it seems like it might rain soon and I do not like to be stuck in the **rain**..."*

2

For every content word (**w**), calculate **incremental lexical probabilities** **P(w_n | w_{n-1} ... w_{n-k})** by increasing the length of the prior context (*k)* by one word each time until reaching 50 words back (*k* = 50).

Example for a single content word: rain

Prior context fed to GPT		***P* (rain | prior context)**
N-1	(the...)	*P*(**rain** | the)
N-2	(in the...)	*P*(**rain** | in the)
N-3	(stuck in the...)	*P*(**rain** | stuck in the)
...		...
N-10	(and I do not like to be stuck in the...)	*P*(**rain** | and I do not like to be stuck in the)
...		...
N-25	(their hands and it seems like it might rain soon and I do not like to be stuck in the...)	*P*(**rain** | their hands, and it... stuck in the)
...		...
N-50	(It looks like the man is sitting on a blanket in the park and it seems like behind him there are dogs running and people talking with leashes in their hands and it seems like it might rain soon and I do not like to be stuck in the...)	*P*(**rain** | It looks like the man... stuck in the)

3

Repeat for all content words in a single sample, from each participant (**35 Healthy Controls** ——, **60 First Episode** ▪▪▪). Then, **compute grand averages** for each context distance, for each group.