

**Taking it with a grain of salt and pepper: Spared and impaired use of contextual
information in the language production of people with schizophrenia**

A dissertation submitted by

Victoria P. Sharpe

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Tufts University

August 2023

© 2023, Victoria P. Sharpe

Advisor: Gina Kuperberg, M.D., Ph.D.

Abstract

Schizophrenia is a severe neurodevelopmental disorder that is characterized by atypicalities in cognition, perception, and psychosocial function. Though these atypicalities are varied, language dysfunction is considered central to the schizophrenia syndrome and has been of interest to researchers and clinicians for many years. However, the mechanisms driving language dysfunction in schizophrenia are poorly understood. One reason for this is that areas of the field that are concerned with the neurocognitive mechanisms underlying language dysfunction have disproportionately focused on language *comprehension* rather than language *production*. Another reason is that experimental studies of language in schizophrenia have been largely limited to single-word paradigms, ignoring the constraints and complexities that bear on natural communication. In the present thesis, I begin to address these gaps in the literature, with emphasis on how people with schizophrenia use different types of context during language production. The findings discussed herein paint a picture of selectively impaired language production in schizophrenia, in which the use of some aspects of context is spared and use of others is atypical. I discuss the conclusions that can be drawn from these findings, as well as their broader implications within the wide-ranging and varied field of schizophrenia research.

Acknowledgments

This research was funded, in part, by the Sidney Baer Foundation. Chapter 3 was adapted from a published manuscript that reflects a collaborative effort by multiple authors (Sharpe, Schoot, Lewandowski, Öngür, Türközer, Hasoğlu , & Kuperberg, 2022). Chapter 4 reflects a collaborative effort with former research assistant Santiago Noriega. Chapter 5 is based on speech data originally collected by Dr. Lena Palaniyappan’s group, to whom I am extremely grateful for sharing this wonderful dataset. Collection of this data was funded, in part, by a Canadian Institutes of Health Research Foundation Grant (154296).

I would also like to thank my committee members: Dr. J.P. de Ruiter, Dr. Lena Palaniyappan, Dr. Eric Tan, Dr. Lin Wang, and Dr. Gina Kuperberg (Chair). Gina also served as my advisor for my six years as a graduate student at Tufts, and I am infinitely grateful for her endless support and guidance.

I am also grateful to the rest of the Kuperberg lab, including a number of undergraduates who helped with the studies reported herein: Brooke Brody, Feng Cheng, Emma Draisin, Zoe Gardaret, Emma Hayden, Julia Klein, Sayaka Koga, Grace Konstantin, Emma Lampropoulos, Rebecca Schaub, and Maggie Wargo. Thanks also to Jaisal Merchant for her critical role in data collection for Chapter 3.

A huge thank you, also, to my family, for their endless encouragement and love. I’m especially thankful for my wonderful partner, Samir—who has supported me continuously in innumerable ways over the past eight years, and particularly over these last few quite difficult months—and for Gene, Burt, Indy, Nandor, Laszlo, Colin, Evie, and Nadja. I couldn’t possibly imagine the last eight years without you all.

TABLE OF CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	vii
CHAPTER 1. GENERAL INTRODUCTION	1
CHAPTER 2. LANGUAGE AND SCHIZOPHRENIA IN CONTEXT: A SELECTIVE LITERATURE REVIEW OF SIX DIFFERENT APPROACHES TO STUDYING LANGUAGE IN SCHIZOPHRENIA	5
CHAPTER 3. WE BOTH SAY TOMATO: INTACT LEXICAL ALIGNMENT IN SCHIZOPHRENIA AND BIPOLAR DISORDER	30
CHAPTER 4. PATTERNS OF LEXICAL PREDICTABILITY IN HEALTHY ADULTS' SPEECH: EVIDENCE FROM GPT-2	51
CHAPTER 5. LEXICAL PREDICTABILITY IN SCHIZOPHRENIA: EVIDENCE FOR SELECTIVE DEFICITS IN USE OF GLOBAL CONTEXT DURING LANGUAGE PRODUCTION	74
CHAPTER 6. GENERAL DISCUSSION	109
APPENDIX A	118
APPENDIX B	139
REFERENCES	143

LIST OF FIGURES

Figure 1. Setup of the cooperative picture naming game. The participant and experimenter faced one another, each looking at their own computer screen, which was not visible to the other player. On the i th trial, the experimenter named a dual-name picture, and the participant, looking at his own screen, then stated whether the picture that he was looking at was the “same” or “different.” After two intervening trials (the $i+3$ th trial), the participant named the same picture (here, using the same name the experimenter used), and the experimenter said “same” or “different.” 36

Figure 2. (A) Probability of using the dispreferred name after hearing the experimenter use the dispreferred name versus the preferred name. Error bars represent one standard error. (B) Lexical production priming: predicted naming times for repeated versus unrepeated low frequency single-name pictures, after controlling for demographic and lexical factors and for mean response time. Error bars represent one standard error. (C) Positive correlation between lexical production priming effects and alignment probabilities within the schizophrenia group. Points represent actual data, whereas the line represents the estimated regression coefficient (Est. = 0.23; SE = 0.09; $p=0.01$) and the shaded region around it represents the predicted standard error. (D) Positive correlation between mentalizing (TASIT scores) and alignment probabilities within the schizophrenia group. Points represent actual data, whereas the line represents the estimated regression coefficient (Est. = 0.32; SE = 0.08; $p=0.00$) and the shaded region around it represents the predicted standard error..... 41

Figure 3. Blue line represents line of best fit based on OLS regression with Cloze Lexical Predictability as the dependent variable. Intercept: .09 (SE = .002, $t = 35.75$, $p < .001$); Slope: .94 (SE = .008, $t = 112.83$, $p < .001$). 61

Figure 4. Blue line represents line of best fit based on OLS regression with Cloze Surprisal as the dependent variable. Intercept: .95 (SE = .018, $t = 51.94$, $p < .001$); Slope: .27 (SE = .003, $t = 78.16$, $p < .001$). 62

Figure 5. Blue line represents line of best fit based on OLS regression with Cloze Constraint as the dependent variable. Intercept: .27 (SE = .003, $t = 78.94$, $p < .001$); Slope: .64 (SE = .009, $t = 71.56$, $p < .001$). 62

Figure 6. Blue line represents line of best fit based on OLS regression with Cloze Entropy as the dependent variable. Intercept: .33 (SE = .006, $t = 54.58$, $p < .001$); Slope: .43 (SE = .004, $t = 88.91$, $p < .001$) (Though, we note that the results of this regression are likely unreliable due to the strong non-linearity of the relationship.)..... 63

Figure 7: Mean Full Context Predictability plotted against Word Position. Lines represent smoothed data via local regression; shaded regions represent 95% CI..... 65

Figure 8. Mean Full Context Predictability plotted against Log Word Position, Lines represent smoothed data via local regression; shaded regions represent 95% CI..... 66

Figure 9. Mean Sentence Context Predictability plotted against Log Word Position, Lines represent smoothed data via local regression; shaded regions represent 95% CI.	67
Figure 10. The effect of Group and Context Type on surprisal as predicted by the LMER. Error bars denote 95% CI.	88
Figure 11. Effect of Group and Context Type on surprisal, as predicted by the LMER. Error bars denote 95% CI.	91
Figure 12. Effect of Group and Log Context Length on surprisal, as predicted by the LMER model. Error bars represent 95%.....	94
Figure 13. The effect of Log Context Length and TLI Disorganization on surprisal. Shaded areas represent 95% CI.	100
Figure 14. The effect of Log Context Length and TLI Impoverishment on surprisal. Shaded areas represent 95% CI.	102
Figure A1. Plot of data by Log Context Length and Group. Lines represent smoothed data using local (“loess”) regression, with span = .5. (Cleveland & Devlin, 1988) Shaded area represents standard error (as estimated by the local regression procedure)	141
Figure A2. Robustness of the negative effect of Context Length on Surprisal within healthy controls. By-subject lines represent smoothed data using local (“loess”) regression, with span = .5. (Cleveland & Devlin, 1988) Shaded area represents a 95% CI, as estimated by the local regression procedure.	142

LIST OF TABLES

Table 1. Means are indicated, with standard deviations in parentheses. ^a M: Male, F: Female. ^b AI: American Indian/Alaska Native, A: Asian, AA : African American/Black, W: White, MR: More than one race, NR: Unknown or not reported. ^c Hawaiian/Pacific Islander was also considered, but no participants identified with this group. ^d SES: Parental Socioeconomic Status, assessed using the Hollingshead Index (Hollingshead, 1965). ^e The Awareness of Social Inference Test (McDonald et al., 2003). ^f PANSS: Positive and Negative Syndrome Scale (Kay et al., 1987). ^g Young Mania Rating Scale (Young et al., 1978). ^h T: Typical, AT: Atypical. ⁱ CPZ: Chlorpromazine equivalent, see Appendix A for details of calculation. The three groups were matched on gender, $X^2(2, 95)=2.484$, $p = .289$, but it was not possible to match all groups on age, parental socioeconomic status, or premorbid verbal IQ. To account for potential confounds, these three variables were covariates in all analyses in which Group was a predictor of interest.	35
Table 2. [^] Marginally significant with slightly greater alignment in the bipolar group. Logistic regression analyses examining the log odds of lexical alignment within and between groups. Effects of predictors of interest are shown. Note that alignment within each group is given by the Intercept of each model, which describes the log odds of the participant's word choice matching that of the experimenter. See Appendix A for full results.	40
Table 3. Effect of Alignment on log-transformed naming times within and between groups. Effects of predictors of interest are shown. See Appendix A for full results. Statistical significance was assessed using a type-III sums of squares estimation, with p-values estimated using the Satterthwaite approximation (Satterthwaite, 1946).	42
Table 4. Effect of Repetition (repeated vs. non-repeated) on log-transformed naming times within and between groups. Effects of predictors of interest are shown. See Appendix A for full results. Statistical significance was assessed using a type-III sums of squares estimation, with p-values estimated using the Satterthwaite approximation (Satterthwaite, 1946).	43
Table 5. Logistic regression analyses examining the effect of mentalizing (TASIT) and lexical production priming (Priming) on the log odds of alignment in patients. Effects of predictors of interest are shown. See Appendix A for full results.	45
Table 6. Effect of Word Position and Speech Segment Length on Full Context Surprisal. Predictors of interest shown in bold.	65
Table 7. Effect of Word Position in Sentence and Sentence Position in Segment on Sentence Context Surprisal. Predictors of interest shown in bold.	68
Table 8. Standard deviations in parentheses; significant differences between groups shown in bold ^a F = Female; M = Male; NB = Non-Binary/Intersex; ^b Digit-Symbol Substitution Test, mean of Written and Oral scores; ^c Category/Semantic Fluency: number of animal exemplars produced; ^d Trail-Making Test, Part B; ^e Mean length of speech segment (in words); ^f Percent of words produced, per utterance, that were content words (vs. function words or disfluencies)	80

Table 9. Effect of Group and Context Type (NoContext vs. RelatedContext) on Surprisal. Predictors of interest are shown in bold.	87
Table 10. Effect of Group and Context Type (NoContext vs. SemanticContext) on Surprisal. Predictors of interest are shown in bold.	90
Table 11. Effect of Group and Log Context Length on Surprisal. Predictors of interest are shown in bold.....	93
Table 12. Effect of Group on Global Surprisal. Predictors of interest are shown in bold.	95
Table 13. Effect of Group on Local Surprisal. Predictors of interest are shown in bold.....	95
Table 14. Effect of Overall Cognition, Group, and Log Context Length on Surprisal. Predictors of interest are shown in bold.	97
Table 15. Effect of TLI Disorganization and Log Context Length on Surprisal (Within Patients). Predictors of interest are shown in bold.	98
Table 16. Effect of TLI Disorganization on Global Surprisal (Within Patients). Predictors of interest are shown in bold.....	99
Table 17. Effect of TLI Disorganization on Local Surprisal (Within Patients). Predictors of interest are shown in bold.....	99
Table 18. Effect of TLI Impoverishment and Log Context Length on Surprisal (Within Patients). Predictors of interest are shown in bold.	101
Table A1. Preferred and dispreferred names of dual-name pictures, used to examine lexical alignment, and low-frequency names of single-name pictures used to examine lexical priming (matched for frequency with dispreferred names).	123
Table A2. Distribution of trials across conditions.	127
Table A3. Effect of Group on Log Odds of Lexical Alignment. Logistic mixed effects regression. Statistically significant predictors of interest are shown in bold.	132
Table A4. Effect of Group on Log Odds of Lexical Alignment. Logistic mixed effects regression—no “nuisance” covariates included in the model. Statistically significant predictors of interest are shown in bold.	132
Table A5. Effect of Group on Log Odds of Lexical Alignment. Bivariate correlations between independent and dependent variables. Pearson’s r.	132
Table A6. Effect of Group on Mentalizing (TASIT Score). Linear regression. Statistically significant predictors of interest are shown in bold.	133

Table A7. Effect of Group on Mentalizing (TASIT Score). Linear regression—no “nuisance” covariates included in the model. Statistically significant predictors of interest are shown in bold.	133
Table A8. Effect of Group on Mentalizing (TASIT Score). Bivariate correlations between independent and dependent variables. Pearson’s r.	133
Table A9. Effect of Group and Repetition on Naming Response Times. Mixed effects linear regression. Statistically significant predictors of interest are shown in bold.	134
Table A10. Effect of Group and Repetition on Naming Response Times. Linear mixed effects regression—no “nuisance” covariates included in the model. Statistically significant predictors of interest are shown in bold.	135
Table A11. Effect of Group and Repetition on Naming Response Times. Bivariate correlations between independent and dependent variables. Pearson’s r.	135
Table A12. Effect of Group and Alignment on Naming Response Times. Within trials preceded by the dispreferred name; mixed effects linear regression. No predictors of interest were statistically significant.	136
Table A13. Effect of Group and Alignment on Naming Response Times. Within trials preceded by the dispreferred name; mixed effects linear regression—no “nuisance” covariates in the model. No predictors of interest were statistically significant.	136
Table A14. Effect of Group and Alignment on Naming Response Times. Bivariate correlations between independent and dependent variables. Pearson’s r.	137
Table A15. Exploratory analysis of the effect of symptoms on alignment; mixed effects logistic regression. No predictors of interest were statistically significant.	138

Chapter 1. General Introduction

Schizophrenia is a severe neurodevelopmental disorder that is characterized by a wide variety of atypicalities across multiple domains of cognition, perception, and psychosocial function. Given the central role that language plays in all of these domains, it is unsurprising that language dysfunction is considered central to the schizophrenia syndrome. Indeed, language has been of interest to schizophrenia researchers ever since the disorder was first characterized.

To date, the study of language in schizophrenia has been approached in a number of ways, ranging from those more concerned with clinical practice - diagnosing and treating patients, and predicting outcomes and progression of symptoms - to those more interested in the neurocognitive underpinnings of the disorder. These varying approaches make for a broad, rich literature with input from multiple fields. However, this literature is also marked by major disconnects between these individual approaches, such that they in many cases seem to run parallel to one another, rather than complementing each other. In Chapter 2 of this thesis, we discuss, broadly, six different approaches that have been taken to studying language in schizophrenia and other psychoses: clinical, computational, neuropsychological, socio-cognitive, psycholinguistic, and neuroimaging. We highlight gaps within each of these domains, with an eye toward developing a more comprehensive, clear picture of how and why language is affected in schizophrenia.

To preview our conclusions, two important themes emerge from Chapter 2 that will motivate the remaining chapters in this thesis. One is that of a great imbalance in how language production versus language comprehension are treated within these domains. Generally, there has been very little clinical interest in language comprehension, despite psycholinguistic and neuroimaging studies showing reliable deficits in how patients comprehend language.

Conversely, the study of the neurocognition of language in schizophrenia has only rarely emphasized language production, despite the evidence, in clinical domains, that dysfunctional language output is central to the disorder *and* predictive of functional outcomes (see Chapter 2 for discussion).

Another theme that has emerged is that of a conspicuous lack of more naturalistic paradigms within those domains concerned with the mechanistic underpinnings of language in schizophrenia - i.e. primarily psycholinguistics and neuroimaging. On the one hand, simple, tightly controlled experimental designs like semantic priming in comprehension and semantic fluency in production have been immensely important to these literatures, identifying areas of obvious impairment that hint at fundamental deficits. On the other hand, using language in the real world to communicate is quite different from engaging in these paradigms, imposing a number of additional constraints — e.g., the need to take one's conversation partner's knowledge into account. If, ultimately, the goal is to be able to understand how language impairments affect patients' communication in the real world, it is clear that it will be necessary to incorporate some of these constraints into our studies, even if that means sacrificing some experimental control.

Chapters 3, 4, and 5 of this thesis are concerned with beginning to address these gaps in the literature, with a focus on understanding the mechanisms underlying disordered language production in schizophrenia. Specifically, we ask how patients use different types of context to produce language. The question of how patients use context has been of great interest in the study of language comprehension, and there have been relatively consistent findings that show selective deficits in using certain types of context to predict and process upcoming words. However, there has been almost no work examining whether and how, use of context varies in language *production* - with the exception of a handful of small studies in the 1970s, which are

discussed in Chapter 5.

In Chapters 3 and 5, we probe language production in schizophrenia under more naturalistic conditions, thus incorporating some of the aforementioned constraints in order to bring our studies closer to how patients might use language in the real world. Chapter 3 describes a study recently published in *Schizophrenia Research*, in which we begin to bridge between psycholinguistics and aspects of social cognition by setting up a task with a partner, and clear cooperative, communicative goals. Here, we probe lexical alignment in schizophrenia within the context of these goals and the partner's speech, juxtaposing top-down use of higher level context (mentalizing) with bottom-up use of low-level context (lexical priming). Though this task is in reality carefully constructed and controlled, the sense is one of naturalism, as a game.

In Chapter 4, we focus on healthy adult subjects, beginning to explore techniques and measures relating to a particular series of state-of-the-art large language models: GPT. In particular, we will use GPT-2 in a series of naturalistic and natural language texts to show that lexical predictability - a major predictor of processing difficulty in language comprehension - exhibits reliable, theoretically-relevant patterns in the the language output of healthy adults. These patterns suggest that healthy adults rely heavily on global linguistic context for producing coherent speech. While these findings are interesting in their own right, this study, more importantly, serves to scaffold the work in Chapter 5.

In Chapter 5, we return to schizophrenia, using GPT-3 to examine naturalistic speech in a sample of patients experiencing their first episode of psychosis, relative to a demographically-matched sample of healthy controls. Here, we use lexical predictability to quantify use of local and global linguistic context and in doing so, test the hypothesis that patients are selectively impaired at using more global context - a finding that is mirrored in the language comprehension

literature and across multiple other perceptual and cognitive domains. We also link use of local versus global context to descriptive characterizations of speech disorganization, bridging between clinical assessments, computational approaches, and cognitive mechanisms of predictive processing.

Chapter 2. Language and schizophrenia in context: A selective literature review of six different approaches to studying language in schizophrenia

In this chapter, we give a broad overview of six different approaches that have been taken to studying language in schizophrenia and other psychoses: clinical, computational, neuropsychological, socio-cognitive, psycholinguistic, and neuroimaging. Our goal is not to provide a comprehensive review of any one of these approaches, but rather to highlight lines of research that have shaped our current understanding of language dysfunction in schizophrenia. For each approach, we highlight gaps in that domain, arguing that the future of research in language and schizophrenia should be concerned with bringing together the knowledge from these domains in order to form a clearer, more holistic picture of language impairment in schizophrenia.

I. The Clinical Perspective

Nearly all of the symptoms associated with schizophrenia can be linked to language. These range from the misperception of one's own internal speech as coming from an external source (auditory verbal hallucinations) to the failure to produce language spontaneously (alogia) to the reduced emotional expression in speech intonation (prosody) that can characterize negative symptoms (DSM-V; American Psychiatric Association, 2013). However, from a clinical perspective, probably the most prominent language disturbance in schizophrenia is “positive thought disorder” — language output that is disorganized and that the comprehender finds difficult to follow. Thought disordered speech is characterized by “tangentiality” and “derailment” — leaps in the message being expressed (Bleuler, 1950; Chaika, 1974; Andreasen,

1979a, 1979b, 1986) , as well as Bleuler’s well-known “loosening of associations” (Bleuler, 1950) – the tendency to produce semantically associated words that are only indirectly related or unrelated to the overall message. Evaluations of thought disorder are standard in clinical assessments of schizophrenia. Indeed, the DSM-V lists “Disorganized Speech” as one of five primary symptoms of the schizophrenia syndrome (American Psychiatric Association, 2013).

For research purposes, the phenomena comprising thought disorder have been catalogued and quantified using various standardized rating scales (e.g., Andreasen, 1979a, 1979b, 1986; Johnston & Holtzman, 1979; Solovay et al., 1986; Marengo et al., 1985, 1986). We know from the use of these scales that thought disorder frequently occurs in conjunction with other positive symptoms (Andreasen, 1979; Harvey, Earle-Boyer, & Wielgus, 1984; Oltmanns et al., 1984), that it is present in approximately 20% - 50% of people with schizophrenia (Breier & Berg, 2003; Andreasen & Black, 2005), both in the acute and chronic phases of the illness (Marengo & Harrow, 1987, 1997), and that it worsens over time (Marengo & Harrow, 1987, 1997; Maeda et al., 2007). We also know that thought disorder occurs in first-degree relatives of people with schizophrenia (Shenton, Solovay, & Holzman, 1989), as well as youths at clinical high-risk for developing psychosis (Solomon, et al., 2011; Perkins & McGlashan, 2015).

Unfortunately, from a clinical perspective, we know far less about language *comprehension* in schizophrenia. This is, perhaps, not so surprising. The incoherent speech that characterizes positive thought disorder will stand out to the average clinician. However, difficulties in language comprehension, for the most part, do not result in tangible measurable outputs to tell us whether or not a patient has fully understood what we have just said. Therefore, comprehension deficits are far less likely to be picked up in standard clinical assessments.

One exception to this is the tendency to interpret language literally when a figurative

interpretation would be more appropriate. For example, a patient asked to “take a chair” (meaning “sit down”), might literally pick up the chair. This “concrete thinking” has long been noted by clinicians (Searles, 1962; Harrow, Tucker, & Adler, 1972), and some psychopathologists have considered it to be a component of thought disorder (e.g. Holzman et al., 1986). Accordingly “concrete thinking” has been included in various clinical rating scales, e.g. in the Bizarre-Idiosyncratic Thinking Rating System (BIZ; Marengo et al., 1986) and the commonly-used Positive and Negative Syndrome Scale (PANSS; Kay, 1987), where it is usually assessed through proverb interpretation tasks (e.g., Sponheim et al., 2003).

II. Neuropsychology

The field of Neuropsychology aims to characterize cognitive performance across a wide range of domains using standardized, well-validated assessment batteries. These assessments can be used to predict functional outcomes beyond what can be gained from looking purely at symptoms. They can also be used as targets for the development of better drugs and rehabilitation programs. The type of neuropsychological assessments that are most commonly used in schizophrenia were originally developed to distinguish impaired and unimpaired performance in patients with non-focal brain injuries (see Prigatano, Pepping, & Klonoff, 1986, for discussion). They probe a number of domains, including semantic memory, episodic memory, working memory, and executive function.

In this subfield of schizophrenia research, “language” itself has not been considered a cognitive domain. Instead, language has usually been studied through the lens of other domains, using tasks that use words. Patients perform abnormally in a variety of different tasks ranging from semantic categorization (e.g., Gold, Randolph, Carpenter, Goldberg, & Weinberger, 1992;

Chen, Wilkins, & McKenna, 1994), word association (e.g., Moran, Mefferd, & Kimble, 1964; Levine, Schild, Kimhi, & Schreiber, 1996), semantic fluency (e.g., Allen & Frith, 1983; see Bokar & Goldberg, 2003; Tan, Neill, Tomlinson, & Rossell, 2020 for review), verbal working memory and episodic memory (recall and recognition of word lists) (Nestor, Akdag, O'Donnell, Niznikiewicz, Law, Shenton et al., 1998; see Bowie & Harvey, 2005 for review). Poor performance in verbal learning and verbal memory tasks are particularly likely, compared to other cognitive measures, to predict poor functional outcomes in schizophrenia (Green, Kern, Braff, & Mintz, 2000).

Another way that language has been studied within this subfield is by examining correlations between performance in various neuropsychological tasks and clinical/offline measures of language function. For example, severity of positive thought disorder has been shown to correlate with performance on a number of cognitive tasks (Docherty & Gordinier, 1999; Kerns & Berenbaum, 2002; Kerns, 2007; Tan & Rossell, 2014), particularly those designed to measure executive functioning (e.g., Kerns & Berenbaum, 2002; Tan & Rossell, 2014). Additionally, studies by Docherty and colleagues have linked impairments in referential communication to poor performance on tasks probing immediate auditory memory, distractibility, working memory, sustained attention, and sequencing abilities (Docherty & Gordinier, 1999; Docherty 2005). Others have shown that deficits in working memory can predict impairments in listening comprehension (Condray et al., 1996; Bagner, Melinder, & Barch, 2003), and that impaired executive functioning can predict the ability to distinguish anomalous and plausible sentences (Kuperberg, 1998; Tan & Rossell, 2019) as well as the ability to correctly comprehend figurative language (Thoma et al., 2009).

There have been far fewer neuropsychological studies that have systematically probed

language beyond the single word in schizophrenia. However, a few researchers have studied schizophrenia using standardized neuropsychological batteries that were designed to study language function in aphasia. These studies have shown that patients exhibit clear impairments in both reading (e.g., Hayes & O’Grady, 2003; Arnott, Sali, & Copland, 2011) and listening (e.g., Goren, Tucker, & Ginsberg, 1996; Pawelczyk et al., 2017) comprehension. Notably, impairments in reading comprehension are very strongly linked to social and occupational functioning (McGurk & Meltzer, 2000; Revheim et al., 2014; Bambini et al., 2016).

Despite these findings, standardized tasks that probe language beyond the single word are not included in batteries designed to provide targets for drugs/treatment in schizophrenia (e.g. the MATRICS; Green & Neuchterlein, 2004). Indeed, patients with language comprehension deficits may even be excluded from these types of studies and tests; for example, Buchanan and colleagues (Buchanan et al., 2011) caution against recruiting thought disordered patients on the grounds that “the presence of marked positive formal thought disorder may (a) compromise the ability to confirm that the participant has adequately understood test directions and (b) interfere with performance on language-based neuropsychological tests.” This lack of comprehensive neuropsychological assessment of language in schizophrenia are in stark contrast to the comprehensive and systematic assessments of naturalistic language processing that are used in neuropsychological disorders such as autism and dementia.

III. Social Cognition

In schizophrenia, the field of social cognition is concerned with the mental operations needed to “perceive, interpret and process information for adaptive social interactions” (Green, Horan, & Lee, 2019). This subfield in schizophrenia research developed around 20 years ago

when researchers began to appreciate that the types of neuropsychological batteries described above may not necessarily provide the best predictors of functional outcome in patients (Green, Kern, Braff & Mintz, 2000). The goal of this subfield was to develop additional standardized batteries designed to tap into domains of psychosocial function that might provide better prognostic markers of psychosocial function.

Given that language is the system that humans evolved in order to accomplish effective social communication, it is somewhat surprising that, similar to standard Neuropsychological approaches, “language” itself is not considered a “domain” of social function in these batteries. However, certain aspects of language comprehension are assessed as part of a group of tasks that aim to probe “mentalizing” (also called “perspective-taking” or sometimes “theory of mind”; Green, Horan, & Lee, 2015)—the ability to model another person’s mind in order to make inferences about their mental state.

First, in the Hinting Task (Corcoran, Mercer, & Frith, 1995), vignettes depicting everyday social interactions are read to/by the patient. Each story ends with a remark that can be interpreted as a hint, e.g. “Rebecca’s birthday is approaching. She says to her Dad, ‘I love animals, especially dogs.’” Patients are then asked to answer comprehension questions about this remark, e.g. “What does Rebecca really mean by this?” Whereas healthy control participants can infer that Rebecca wants her dad to get her a puppy for her birthday, people with schizophrenia often fail to get the “hint”, reflecting difficulties in inferring the true intent of indirect speech (Corcoran et al., 1995; see also Bora, Yucel, & Pantelis, 2009 for review).

Second, in the The Awareness of Social Inference Test (TASIT; McDonald et al., 2003), participants are asked to detect lies and sarcasm in a series of video vignettes. For example, in one vignette, a husband tells his wife that he needs to cancel their dinner date in order to work

late. The wife rolls her eyes and says “No, that’s *fine*.” Here, the utterance should be taken to mean the opposite of its literal meaning. While this will be obvious to healthy adults based on context, people with schizophrenia tend to exhibit impaired performance on the TASIT (e.g. Kern et al., 2009; Roberts & Penn, 2009; Sparks et al., 2010) , which has been linked to poorer recreational social functioning (Sparks et al., 2010).

Both these tasks rely on patients being able to use an aspect of language function known as “pragmatics”. Pragmatics describes the general principles that healthy individuals implicitly adhere to in order to facilitate everyday interactions. They are thought to be well-described by Grice’s famous maxims (see examples below; Grice, 1975) as well as theoretical frameworks like Relevance Theory (Sperber & Wilson, 1995; Wilson & Sperber, 2002), which states that speakers strive to communicate information which is pertinent to the topic at hand, and expect their partner’s speech to meet certain standards of relevance. For example, returning to the above example from the Hinting Task, whereas Rebecca’s remark might appear irrelevant at face value, its presumed link to the context (her upcoming birthday) allows healthy adults to infer what Rebecca actually meant. Similarly, as in the TASIT, detecting sarcasm requires participants to identify situations in which speakers *intentionally* flout the Gricean maxim of quality (which dictates that speakers speak truthfully to the best of their ability). Indeed, a failure to detect such violations may explain why people with schizophrenia are less able than healthy controls to detect verbal irony (Mitchley et al., 1998; Langdon, Davies, & Coltheart, 2002) more generally, or to comprehend verbal humor (Bambini et al., 2016; Pawelczyk et al., 2017), given that these types of figurative language are often achieved by flouting pragmatic principles.

These findings therefore suggest some systematic problems with use of pragmatics in schizophrenia. One problem, however, is that, because these tasks are generally grouped together

with non-linguistic tasks, e.g. interpreting facial expressions, it is not altogether clear the degree to which language-specific mentalizing mechanisms may be affected. In addition, performance in the tasks like the TASIT does not rely purely on linguistic information, but also on a combination of prosody, gesture, and visual context.

Finally, and perhaps most importantly, social pragmatic principles that pertain to mentalizing are not simply relevant in comprehension. They also play a key role in production. Indeed, many aspects of thought disorder can be viewed as a failure to take the partner's perspective into account (Rochester, Thurston, & Rupps, 1977; Hoffman et al., 1982; Hoffman, 1986, see also Rutter, 1985; Harrow, Lanin-Kettering, & Miller, 1989). Importantly, however, social communication through language is rarely one-sided; it requires to-and-fro communication between interlocutors. This sort of back-and-forth is dependent on continuous mentalizing to facilitate language comprehension and production as a joint enterprise. Though there are some assessments that involve back-and-forth communication via role-playing scenarios and natural-like dialogues—for example, the Social Skills Performance Test (Patterson et al., 2001; Roberts & Penn, 2009) and the Schizophrenia Communication Disorder Scale (Bazin, Sarfati, Lefrere, Passerieux, and Hardy-Bayle, 2005)—there has been little systematic study of how mentalizing deficits shape patients' ability to engage in natural-like to-and-fro conversation. Thus, many gaps remain in our knowledge of how social cognition is borne out through language in people with schizophrenia.

IV. Descriptive Computational Approaches

Over the past decade, a very different approach to characterizing language production in schizophrenia has emerged: The use of *computational language models* to quantitatively describe

aspects of patients’ natural language output. Taking advantage of modern advances in computational power and powerful machine learning methods, these computational models are typically “trained” on large corpora of natural language outputs such that they “learn” multiple types of statistical regularities within language. Then, these normative statistical regularities can be used to generate various descriptive measures of language produced by patients.

Multiple different types of measures have been used to describe language output in schizophrenia. For example, a number of semantic similarity measures have been developed using vector space models like LSA and word2vec, which encode word co-occurrence as numerical vectors to approximate word meaning (e.g., Elvevag et al., 2007; Bedi et al., 2015; Corcoran et al., 2018; Iter & Jurafsky, 2018). There has also been interest in graph-theoretic “connectedness” measures, which rely on links between words or phrases extracted from “speech graphs” (e.g., Mota et al., 2012, 2016, 2017; Palaniyappan et al., 2019; Spencer et al., 2021). Another approach has focused on estimation of referential ambiguity (number of ambiguous pronouns) using neural-network-based co-reference resolvers (e.g., Iter & Jurafsky, 2018). Finally, a handful of studies have also begun to employ language models to generate estimates of lexical predictability in the speech of people with schizophrenia, though so far with limited success (Mitchell, Hollingshead, & Coppersmith, 2015; Cecchi et al., 2019). In Chapters 4 and 5 of this thesis, we discuss how large language models like GPT (Radford et al., 2019) may allow us to assess lexical predictability in schizophrenia more systematically.

Most of these studies have emphasized the *discriminatory* power of these computational measures—that is, whether they are able to accurately *classify* participants into patients *versus* control groups (Elvevag et al., 2007; Mota et al., 2012, 2017; Mitchell et al., 2015; Corcoran et al., 2018; Iter & Jurafsky, 2018; Voleti et al., 2019). Certain computational measures have also

been used to predict development of psychosis in clinical high-risk individuals (Bedi et al., 2015; Corcoran et al., 2018; Spencer et al., 2021). In addition, some measures have been shown to correlate with subjective ratings of thought disorder (Elvevag et al., 2007; Iter & Jurafsky, 2018; Palaniyappan et al., 2019), negative symptoms (Iter & Jurafsky, 2018; Mota et al., 2016), and global well-being (Arevian et al., 2020).

Given the relative ease of collecting natural speech samples from patients, and the speed, automaticity, and relative objectivity of computational methods, it has been suggested that these measures may provide a *biomarker* for psychosis that could facilitate earlier detection of illness, symptom monitoring, prediction of outcomes, and other types of clinical decision-making (see Corcoran et al., 2020 for discussion). In addition, the recent push for longitudinal studies which capture different stages of disease progression means that these language assessments may have a role in refining our understanding of the trajectory of language dysfunction in psychosis.

Despite the obvious promise of computational methods, it is important to emphasize that these methods are still largely in their infancy, and many open questions remain. First, the emphasis on the discriminatory capabilities of these measures, for use in classification and predictive modeling, means there has been relatively less emphasis on precisely how these measures relate to thought disorder as a clinical construct. This is important because, even though some measures may have face validity with specific dimensions of positive thought disorder, they do not necessarily map cleanly onto traditional thought disorder rating scales (though see Elvevag et al., 2007).

Second, it is important to recognize that there are many different types of language models, and many different ways to train a given model. Using different parameters or different training corpora can produce different results. It is not well-understood how these differences

affect the construct validity of the resultant measures. Relatedly, there has been little emphasis on the degree to which a given set of findings can *generalize* across additional cohorts of patients, and/or even within the same patients across different types of speech elicitation protocols (e.g. picture description *vs.* clinical interview, although see Corcoran et al., 2018).

More generally, we note that these computational approaches are primarily *descriptive* in nature; they are data-driven with little theoretical basis. Therefore, as for any descriptive variables, the interpretation of such measures depends largely on how they are used: what questions the researcher wants to address, the specific statistical analyses conducted, etc. Indeed, there are thousands of dimensions on which language output can be quantified, and many of them are likely to be predictive of psychosis. However, using computational language models to their full potential in schizophrenia will require developing a deeper understanding of how these measures relate to broader theoretical frameworks.

V. Psycholinguistics

Psycholinguistics is the study of how the brain draws upon the linguistic code (semantics, syntax, phonology, orthography, etc.) to communicate. Because speakers and comprehenders must use this code rapidly as language unfolds in real time, psycholinguistics generally employs methods and tools that have high temporal resolution. These include behavioral methods, in which researchers might analyze response times or eye movements. It also includes neuroimaging methods like event-related potentials (ERPs) — a direct online measure of brain activity that is detected at the scalp’s surface (see also magnetoencephalography, MEG, discussed later). In particular, an ERP component known as the N400, which has been shown to be a sensitive index of lexicosemantic processing, is frequently used in the psycholinguistics

literature (see Kutas & Federmeier, 2011, for review).

Using these tools, a large body of psycholinguistic research has established that, in healthy individuals, when a linguistic input (e.g. a word) is encountered (here referred to as a “target” input), its processing critically depends on its prior *context*. This context may be a single word (such as in priming paradigms), a sentence frame, or a whole discourse. No matter the type of context, targets that can be predicted based on this prior context are “easier” to process, i.e. are facilitated, relative to those that cannot be predicted based on the context. This facilitation manifests behaviorally as shorter processing times (button presses or eye movements) and electrophysiologically as reduced amplitude of event-related potentials, such as the smaller N400 produced by words that are semantically expected versus unexpected. Very broadly speaking, these facilitation effects are thought to be governed by two types of mechanisms.

The first of these mechanisms involves a relatively “passive” and indiscriminate “spread of activation” across linguistic information stored in long-term memory (the “lexicon”). This stored linguistic information can be metaphorically described as a network of internal linguistic representations, which encapsulates our probabilistic knowledge of the statistical regularities that are encoded within linguistic inputs. These include both lower-order correlational relationships based on co-occurrence (e.g. the phoneme /p/ is often followed by the phoneme /r/ but rarely is followed by /t/) as well as more complex higher-level correlations (e.g. determiners like “the” or “a” can be followed by a noun but not by a verb). Prior linguistic observations — the linguistic context—act as cues that activate this network such that activation spreads implicitly to predicted upcoming content. If the target input overlaps with this newly activated content, then the target is easier to “access” —that is, its processing is facilitated. This implicit spread of activation can be thought of as a type of pattern completion.

On the one hand, this type of facilitation can describe activation which spreads *within* a particular level of representation. For example, the word “cheese” and the word “crackers” are related semantically; that is, they share semantic features and are often encountered together. Spread of activation within a level of representation is typically probed using “priming paradigms” under “automatic” experimental conditions, which might include a short interval between prime and target or the use/avoidance of specific tasks (Neely, 1977; Neely & Keefe, 1989). Here, the facilitation effect is called an automatic priming effect. During natural language comprehension, however, activation spreads across the language network not only within levels of representations, but also interactively *across* levels of linguistic representation. These different levels of representation are encoded at multiple time scales and thus the network can be conceptualized as being organized hierarchically, with activation flowing up and down the hierarchy. Importantly, this must happen incrementally as each input becomes available in real time in order to build event representations (who does what to whom) at the top of the hierarchy (see Kuperberg, 2022, for discussion). Thus, comprehenders incrementally process each incoming word as the input unfolds over time.

The second type of mechanism involves the ability to more actively modulate activity within the linguistic network in ways that are relevant to one’s overarching goals. Here, comprehenders draw upon a goal-relevant representation that interfaces between the linguistic network and the non-linguistic representations. Because this goal-relevant representation must be maintained over a longer timescale, it is often referred to as a “higher-level” representation. Thus, information can be conceptualized as flowing “downwards” to enhance activation within the parts of the linguistic network that are most relevant to the goals at hand. Thus, goal-relevant representations receive more facilitation than they would simply through spread of activation.

We refer to this flow of information as *top-down goal-relevant mechanisms*.

Top-down goal-relevant mechanisms can also be studied using priming paradigms, but this time under “non-automatic” experimental conditions (e.g. a longer time interval between prime and target and certain tasks, like lexical decision) that encourage participants to set up and use a higher-level goal-relevant representation (Neely & Keefe, 1989; Hutchison, 2007). The non-automatic semantic priming literature has established that top-down goal-relevant mechanisms can be either strategic (Becker, 1980; Neely, Keefe, & Ross, 1989) or implicit (Lau, Holcomb, & Kuperberg, 2013), and that they can be retroactive (influencing decision-making after the target has been accessed; Neely, Keefe, & Ross, 1989) or proactive (operating predictively to enhance the pre-activation of a target’s representation before the target is encountered; Neely, Keefe, & Ross, 1989).

During natural language comprehension, implicit and proactive top-down goal-relevant mechanisms are thought to play a particularly important role in fast and efficient language comprehension (Kuperberg, 2021). Here, the higher-level goal-relevant representation corresponds to an ongoing *high-level interpretation* of the input. In addition to interacting with the linguistic network, this interpretation interacts with (a) stored non-linguistic real-world schema-relevant knowledge (e.g. the tendency, in some cultures, for competition winners to receive medals or trophies), (b) non-linguistic representations that describe the broader surroundings (e.g. the visual space shared by the communicators, often called the “visual world”), and (c) representations that describe the communicator’s use of language (e.g. knowing that one’s conversation partner speaks a particular dialect). This goal-relevant situation model is continually updated based on the incoming linguistic input, and continually generates top-down predictions that provide an additional boost of activation to representations which correspond to

expected inputs, leading to additional facilitation that goes beyond that achieved only by passive spreading activation.

In schizophrenia, the majority of psycholinguistic studies that have examined patients' use of context during online language processing have used the semantic priming paradigm, in which the context is a single word. This literature provides clear evidence for a dissociation in patients' use of the two types of mechanisms described above. Under relatively automatic experimental conditions, semantic priming effects are generally of the same magnitude in people with schizophrenia as in healthy control participants (e.g. Barch et al., 1996; Blum & Freides, 1995; Chapin et al., 1992; Ober et al., 1995; Vinogradov et al., 1992). This suggests that the passive spread of activity across semantically linked representations is normal in schizophrenia.

In contrast, under non-automatic experimental conditions, the magnitude of the behavioral (Barch et al., 1996; Koyama et al., 1994; Henik, Nissimov, & Umansky, 1995; Goldberg et al., 2000; Moritz et al., 2001a, b) and N400 (Condray, Steinhauer, Cohen, & van Kammen, 1999; Kiang, Kutas, Light, & Braff, 2008; Kreher, Goff, & Kuperberg, 2009) priming effect is smaller in patients than in healthy controls. Of particular relevance to natural language processing is a recent study from our lab in which we showed that, in contrast to healthy adults (Tweedy & Lapinski, 1981; Neely, 1991; Lau, Holcomb & Kuperberg, 2013; Delaney-Busch, Morgan, Lau & Kuperberg, 2019), patients failed to engage *implicit* and *predictive top-down* goal-relevant to enhance facilitate of associated targets when the predictive validity of the experimental environment was high (Sharpe, Weber, & Kuperberg, 2020) .

Far fewer studies have examined the effects of sentence and discourse context on the processing of incoming words in schizophrenia. However, there is some evidence that here, too, there may be a dissociation between preserved spreading activation across the linguistic network,

and impaired use of the top-down goal-relevant mechanisms. When semantically anomalous words are contrasted with highly expected words (“...coffee with milk and dog/sugar”), the magnitude of the N400 effect is generally the same in people with schizophrenia as in healthy adults (Andrews et al. 1993; Kuperberg et al. 2006c; Nestor et al. 1997; Niznikiewicz et al. 1997; Ruchow et al. 2003). This suggests that patients can use some aspects of prior context and that spreading activation is likely preserved during sentence comprehension (although it is unclear from this type of design whether activation spreads only across semantically associated representations, or across both syntactic and semantic representations).

In contrast, people with schizophrenia are less sensitive to prior context in the types of situations in which they cannot rely purely on spread of activity across linguistic network but are, rather, required to proactively mobilize top-down goal-relevant mechanisms to interact with this network. One such situation is the case of ambiguity resolution, such as in homographs (words with two meanings). Indeed, there is a long tradition of studying homographs in schizophrenia (Benjamin & Watt, 1969; Blaney, 1974). In natural language, healthy adults can use real-world knowledge to enhance representations that are consistent with a homograph’s intended meaning, and suppress representations that are consistent with the homograph’s other, inappropriate meaning. For example, correctly comprehending a sentence like “The guests played bridge because the...” requires the comprehender to activate the appropriate meaning of “bridge” (card game) and correspondingly suppress its dominant meaning (a structure designed to cross a gap or obstacle). However, there is evidence that patients fail to suppress non-dominant meanings during online processing of sentences and so show, paradoxically, inappropriate facilitation of information consistent with the homograph’s unintended meaning (e.g., in the above example, facilitated processing of “river” in the sentence continuation “because the river had rocks in it.”)

(Titone, Levy, & Holzman, 2000; Sitnikova et al., 2002).

Lexical ambiguity is, of course, only one type of linguistic ambiguity. In natural language, ambiguity can occur at all levels of representation, even in syntactic structure. Indeed, additional evidence for patients' failure to use top-down goal-relevant mechanisms comes from processing of syntactically ambiguous instructions. Consider, for example, a scene with two toy frogs, one of which is holding a feather, and, beside them, a feather on its own. In a visual world paradigm, the subject might view this scene and be asked to "Tickle the frog with the feather." This instruction is syntactically ambiguous (tickle the frog *using* the feather or tickle the frog *that is holding* the feather?). However, healthy adults can generally use visual context to resolve the ambiguity (the presence of two different frogs suggests that the instruction means to use one's finger to tickle the frog who is holding the feather). Rabagliati and colleagues (2018) found that people with schizophrenia were less able than healthy controls to use visual context to resolve syntactically ambiguous instructions like these in real time, as indexed by their eye movements.

Similarly, during discourse comprehension, one cannot rely on spread of activity alone to build up event representations. Healthy adults are able to rapidly access and mobilize schemas stored in long-term memory such that they inform this event-level interpretation, and, in doing so, mobilize top-down goal-relevant mechanisms to predict schema-consistent information. Consider this example: "James was practicing piano for months. He won first prize in the competition. He took the...". Here, healthy adults can implicitly use their real-world knowledge about competitions to generate appropriate top-down predictions about what should come next. Therefore, when the comprehender encounters an expected word (e.g. "medal"), the N400 is smaller than if the incoming word was unexpected given the prior context (e.g. "ball"). In

contrast, in people with schizophrenia, the N400 is relatively insensitive to these broader discourse constraints (Ditman & Kuperberg, 2007; Swaab et al., 2013; Boudewyn et al., 2017).

One question these studies raise is whether the use of high-level context in schizophrenia is selectively *proactively* impaired, with intact reactive mechanisms. Indeed, there is some evidence that, while proactive processes may be impaired, given enough time, patients *can* engage slower reactive mechanisms to compensate. For example, in the visual eye tracking study by Rabagliati and colleagues (2018), though eye movements suggested that predictive use of context was impaired in schizophrenia, patients were able to disambiguate the instructions eventually, as evidenced by their final actions (e.g. what they did with the frog). Similarly, in a study by Kuperberg, Ditman, and Choi-Perrachione (2018), though patients failed to produce an Nref effect (an ERP that is larger in response to ambiguous, versus unambiguous referents, e.g. “Mark and John went to the store, and *he*...”) to referential ambiguity, they did produce a late positivity effect that has been associated with more retroactive analysis (see also Ditman & Kuperberg, 2007, for similar findings).

Given the relatively few studies of sentence and discourse processing in schizophrenia, major gaps in our knowledge remain. In particular, the dissociation between spreading activation and top-down mechanisms has not been studied systematically in sentence/discourse comprehension as it has been in semantic priming. This underlines the need for more careful examination of patients’ impairments in using sentence/discourse contexts to comprehend incoming information.

Moreover, there is a conspicuous lack of research on the mechanisms underlying impaired language *production* in schizophrenia. There are, however, several studies which have examined production priming in schizophrenia, though this is generally framed in terms of

implicit memory rather than language processing per se (Soler et al., 2014; Marques et al., 2016; though see also Kuperberg et al., 2017). Additionally, some experiments have evaluated aspects of negative thought disorder in perturbed versus non-perturbed processing capacity (Barch & Berenbaum, 1994) and discourse planning (Barch & Berenbaum, 1997). However, there has been very little emphasis on systematically probing the cognitive underpinnings of disorganized language production, particularly under more naturalistic conditions.

VI. Neuroimaging

Language processing is supported by multiple neural circuits, localized across the temporal, inferior frontal, and inferior parietal cortices, particularly in the left hemisphere (Hickok & Poeppel, 2007; Kuperberg, 2007; Baggio & Hagoort, 2011; Friederici, 2012; Price, 2012; Hagoort, 2014). These regions are connected structurally through large white matter tracts, including the arcuate fasciculus, the superior longitudinal fasciculus, and the uncinate fasciculus (Catani et al., 2005; Friederici, 2009; Friederici & Gierhan, 2013; Hagoort, 2019). In addition, *functional* connectivity—the degree to which activity in different parts of the brain is synchronized, or coupled—between regions of the language network is critical for normal language processing (Hickock & Poeppel, 2000; Scott & Johnsrude, 2003; Friederici, 2011; Klingbeil et al., 2019).

There is a large body of work that has shown clear structural deficits in the brains of people with schizophrenia. This includes morphometric studies, which have found reduced volume of the superior temporal gyrus in patients relative to controls (see Sun et al., 2009 for review), particularly in relationship to thought disorder (Shenton et al., 1992; Menon et al., 1995); voxel-based morphometry studies, which have found that patients exhibit widespread

reductions in gray matter volume (see Honea, Crow, Passingham, & McKay et al., 2005, for review), and studies using cortical surface-based methods which found that patients exhibit increased thinning of cortical gray matter in the temporal-prefrontal cortex, particularly on the left side (Kuperberg et al., 2003; Wisco et al., 2007; Rimol et al., 2010). In addition, several diffusion tensor imaging (DTI) studies report structural degradation of white matter connections between the left frontal and temporal lobes, although there has been much variability in the literature regarding which specific tracts are affected (see Fitzsimmons, 2013 and Wheeler & Voineskos, 2014, for review).

A number of different approaches have been used to study *functional* language abnormalities in schizophrenia as well. Initially, many of these studies took a “symptom-capture” approach, comparing brain activity within patients when they were and were not experiencing a particular symptom. One early finding was that, when participants spoke freely in the scanner, greater severity of positive thought disorder was associated with reduced activity in some regions—left inferior frontal cortices (McGuire et al., 1998) and left superior temporal cortex (McGuire et al., 1998; Kircher et al., 2001)—but increased activity in other regions—within medial and ventral regions of the temporal lobe, including the anterior fusiform/parahippocampal cortices (McGuire et al., 1998).

A number of other studies have examined brain function in response to specific tasks. Mirroring the neuropsychological approach, most of these studies used single-word tasks like verbal fluency (e.g., Frith et al., 1995; Curtis et al., 1998), semantic encoding (e.g. Kubicki et al., 2003; Ragland et al., 2004), and semantic retrieval (e.g., Ragland et al., 2004; Ragland et al., 2008). These studies typically have not shown overall reduced activity across fronto-temporal cortices but rather, different patterns of activation across regions, with activity in some regions

increased and in others, decreased. For example, in several studies, patients exhibit *decreased* activity within frontal regions of the language network, but *increased* activity within temporal regions (e.g. Ragland et al., 2004). It has been hypothesized that such a pattern may arise from disconnection across regions, in line with Friston's (1998) characterization of schizophrenia as a "disconnection syndrome". In support of this hypothesis, several studies have found evidence for decreased functional connectivity between inferior frontal and temporal cortices during various tasks, including verbal fluency (Frith et al., 1995), word categorization (Frith et al., 1995), word repetition (Frith et al., 1995), word encoding (Wolf et al., 2007), lexical decision (Li et al., 2010), one-back string matching (Griego et al., 2008), and semantic matching (Woodward et al., 2015).

Abnormal functional connectivity has also been observed during "resting-state" in people with schizophrenia. In healthy individuals, it is established that, at rest, ongoing low frequency (<0.1 Hz) patterns of BOLD activity are observed within various networks (see Fox & Raichle, 2007 for review). This includes the left-lateralized temporal-inferior frontal language network associated with language (Binder, Frost, & Hammeke, 1999; see Smitha et al., 2017 for review), with connections across the left temporal and prefrontal lobes (Tomasi & Volkow, 2012). Several studies have shown that patients exhibit abnormal resting-state functional connectivity, particularly within the language network (e.g. Garrity et al., 2007; Hinkley et al., 2010; Liemburg et al., 2012).

A few studies have examined the neural effects of single-word contexts on lexicosemantic processing—i.e. semantic priming. One interesting finding seen in several studies is that a "reverse" priming effect is observed in patients under non-automatic experimental conditions (e.g. long interval between prime and target). That is, whereas healthy adults exhibit less activity to associated vs unrelated words, patients show the opposite pattern, particularly in

regions within ventromedial temporal lobe (fusiform cortex; Kuperberg et al., 2007; Han et al., 2007; Wilson, Rojas, Shatti, & Tregellas, 2013). A similar reverse effect has also been described in people with high familial risk for schizophrenia (Thermenos et al., 2013).

It is possible that this enhanced activity to semantically related words is driven by patients' prolonged top-down efforts to retroactively link indirectly related primes and targets (Kuperberg et al., 2007). Some indirect evidence for this interpretation comes from the finding that this abnormality disappears under highly automatic experimental conditions—i.e. when the prime is masked and the experiment uses a very short interval between prime and target (e.g. 140ms). Indeed, in this situation, in contrast to healthy controls who typically show no evidence of automatic priming for indirectly related targets, patients exhibit greater activity within the temporal fusiform cortex in response to indirectly related vs. unrelated targets, mirroring the N400 effect under these highly automatic conditions (Kuperberg et al., 2019).

Context *beyond* the single word is rarely examined in the neuroimaging literature (though see Kuperberg et al., 2008, who found that patients, unlike controls, failed to recruit the DLPFC, medial frontal and parietal cortices to incongruous vs. congruous sentences). A few studies have explored hypotheses related to hemispheric lateralization in people with schizophrenia (see Crow, 1997), particularly associated with thought disorder (e.g. when completing semantically constraining sentence stems (Kircher et al., 2001) and during metaphor comprehension (Kircher et al., 2014; Mitchell & Crow, 2005). However, these studies are fairly preliminary, and the topic has not yet been explored systematically.

One reason for this may be limitations in the current methodologies. Though neuroimaging methods like fMRI and PET have excellent spatial resolution, their poor temporal resolution means they cannot tell us about the timecourse with which language processing

unfolds in the brain as linguistic information becomes incrementally available. Indeed, fMRI is relatively blind to transient responses that are associated with initial feedforward stimulus-driven activity (Wang et al., 2022) and therefore does not pick up, for example, the N400 (Geukes et al., 2013; Lau, Phillips, & Poeppel, 2008). In future work, the use of magnetoencephalography (MEG), which has good spatial *and* temporal resolution, may provide insights into the timecourse of neural activity as information flows interactively up and down the fronto-temporal hierarchy during real-time language processing.

Conclusions

What has emerged from the literature is a relatively spotty, non-comprehensive picture of how language is affected in schizophrenia compared to other cognitive deficits. In particular, two major gaps in the literature are apparent. First, there is a significant imbalance in how language production versus language comprehension are treated within these different domains. In those approaches more concerned with clinical applications, there has, generally, been inordinate focus on impairments in language production, with little acknowledgement of the comprehension deficits that have been observed under other approaches. Conversely, in domains which have focused on understanding *mechanism* in language impairment, the vast majority of studies have been in language comprehension, despite evidence from clinical data that atypical language output is extremely prevalent in schizophrenia and, importantly, predictive of functional outcomes.

Within clinical domains, that language processing is rarely systematically assessed in people with schizophrenia is likely to contribute to vast underestimation of less clinically obvious deficits in language comprehension. This may in turn limit the predictive power of

symptom assessments, and the precision with which treatments can be targeted. On the other hand, ignoring language production in neuroimaging and psycholinguistic approaches limits the knowledge with which we can connect language impairment to more fundamental deficits in schizophrenia. Understanding the mechanisms underlying language more clearly could open up new paths for developing targeted therapies and psychosocial rehabilitation programs. Therefore, it is to our benefit to bridge across approaches to begin to correct these imbalances.

A second theme that emerges from this literature review is that within more mechanism-focused domains, there is a clear tendency toward tightly-controlled but highly un-naturalistic experimental paradigms, such as semantic priming. Such paradigms, of course, have been immensely important to our understanding of language in schizophrenia thus far and have begun to identify areas of obvious impairment which hint at more fundamental deficits in schizophrenia. However, it is not easy to piece apart the degree to which we can extrapolate the results of such paradigms to the real world, in which communication requires juggling more, and different types of information simultaneously.

It is worthwhile to note that the lack of more naturalistic paradigms is not only at issue in schizophrenia research - researchers in the social sciences have been calling for studies with more external/environmental validity for years. However, this may be particularly important when dealing with patient populations, for whom impairments in language comprehension and production may result in noticeable differences in quality of life (Bambini et al., 2016).

The themes we have highlighted above motivate the remainder of this thesis, in which we begin to fill some of these gaps by targeting semi-naturalistic and naturalistic language production in people with schizophrenia. Toward this goal, we bridge across several of the domains described above by incorporating techniques from psycholinguistics, social cognition,

and computational approaches. In doing so, we are able to draw novel conclusions about the mechanistic underpinnings of impaired communication in schizophrenia.

Chapter 3. We both say tomato: Intact lexical alignment in schizophrenia and bipolar disorder¹

For most healthy adults, conversation is a powerful means of communication. In people with schizophrenia and related disorders, however, communication impairments can negatively impact social interactions and quality of life (e.g., Bowie & Harvey, 2008; Bowie et al., 2011; Buck & Penn, 2015). The present study investigated the cognitive basis of a specific aspect of linguistic communication—*lexical alignment*—in people with schizophrenia and bipolar disorder.

Lexical alignment plays a crucial role in facilitating successful, rewarding communication (Pickering & Garrod, 2006; Fusaroli et al., 2012; Reitter & Moore, 2014). It describes the tendency of speakers to use the same words as their conversational partners to refer to objects in their common environment (Garrod & Anderson, 1987; Brennan & Clark, 1996; Branigan et al., 2011). For example, if during a conversation between Tori and Gina, Tori points out a cute “bunny”, it is more likely that Gina will later refer to this animal as a “bunny” than as a “rabbit”.

Lexical alignment is thought to be largely mediated by bottom-up priming mechanisms in which a “prime” word directly activates its internal lexical representation, facilitating subsequent processing of this word. That is, hearing a particular word spoken by a conversational partner would pre-activate its lexical representation, and so a speaker would subsequently find it easier to retrieve/access this word than if it had not been encountered (Pickering & Garrod,

¹ This study was done in close collaboration with others, particularly Dr. Lotte Schoot. Findings have been published in *Schizophrenia Research*: Sharpe, V., Schoot, L., Lewandowski, K. E., Öngür, D., Türközer, H. B., Hasoğlu, T., & Kuperberg, G. R. (2022). We both say tomato: Intact lexical alignment in schizophrenia and bipolar disorder. *Schizophrenia Research*, 243, 138-146.

2004). Importantly, however, as for other forms of lexical priming (e.g. Posner & Snyder, 1975; Tweedy et al., 1977; Neely, 1991; Lau et al., 2013), this type of bottom-up lexical facilitation can be modulated by higher-level beliefs or goals in a top-down fashion (see Appendix A for a more in-depth discussion of bottom-up versus top-down processes in priming and lexical alignment). In the case of lexical alignment, these beliefs include the speaker’s knowledge about her conversational partner, and her implicit goal of ensuring effective communication (Branigan et al., 2011). Thus, successful alignment is closely related to perspective-taking or *mentalizing*² (Giles et al., 1991) and to *audience design* (the adaptation of our speech in accordance with what we believe our conversational partner can understand; Clark & Murphy, 1982; Bell, 1984). These principles are thought to play an important role in allowing speakers to align to their partners’ use of “dispreferred” names (Brennan and Clark, 1996). For example, returning to the conversation above, Gina’s belief that Tori would understand her better by referring to the animal as a “bunny” may help override her default tendency to use the more frequent name (“rabbit”).

Given the importance of lexical alignment in communication, it is important to understand how it is affected in people with severe mental illness. Schizophrenia, for example, is characterized by impairments across multiple cognitive and social domains (see Green et al. 2019 for review), which can affect communication and impair psychosocial functioning (e.g., Hogarty & Flesher, 1999; Patterson et al., 2001; Bowie & Harvey, 2008; Buck & Penn, 2015; Vaskinn & Horan, 2020). However, most studies of lexical processing in schizophrenia have been carried out in non-communicative environments (for reviews, see Kuperberg 2010a, b; Brown & Kuperberg, 2015).

It is unclear from these previous studies whether one might expect lexical alignment to

² Throughout the paper, we use the terms *mentalizing* and *perspective-taking* interchangeably, and avoid using the term, *theory of mind*. For discussion of this choice, and how these abilities relate to language, see Appendix A.

be spared or impaired in schizophrenia; either would have interesting theoretical implications. On the one hand, studies of lexico-semantic priming suggest that bottom-up influences on priming are preserved (e.g., Vinogradov et al., 1992; Barch et al., 1996) , or even enhanced (e.g., Spitzer et al., 1994; Kreher et al., 2008) in this population (although these effects have been studied in comprehension rather than production). This might lead one to predict that lexical alignment should be intact. On the other hand, people with schizophrenia tend to perform poorly at mentalizing (Brune, 2005). Moreover, lexical priming in schizophrenia is usually reduced (relative to healthy controls) when experimental conditions encourage top-down influence of higher-level goals (Barch et al., 1996; Kreher et al., 2009; Sharpe et al., 2020). These findings might lead one to predict that lexical alignment should be impaired in schizophrenia.

The study of lexical alignment therefore provides a unique opportunity to understand how two fundamental constructs—bottom-up priming and top-down mentalizing—contribute to social-communicative interactions in schizophrenia. It also offers the opportunity to understand how these constructs vary dimensionally *across* diagnostic categories (Insel et al., 2010). For example, there is some evidence that patients with bipolar disorder also exhibit impairments in social and communicative abilities, albeit to a lesser degree than those seen in schizophrenia (Lee et al., 2013; Lee et al., 2017; see also Bora & Pantelis, 2016 and Samame, 2013 for review). However, there has been little work on linguistic communication in bipolar disorder, and *no* studies of lexical alignment in bipolar.

The Present Study

The first goal of this study was to determine whether people with schizophrenia and bipolar disorder can lexically align to their conversational partner to the same degree as healthy adults.

To this end, we probed lexical alignment as participants played a collaborative picture-naming game with the experimenter. This game was based on a paradigm developed by Branigan and colleagues (Branigan et al., 2016) and was designed to mimic the cooperative nature of conversation while eliciting speech in a controlled fashion. The participant and the experimenter alternated in naming pictures that appeared in front of them on separate screens (see Fig. 1). Critically, a subset of pictures had two names—a preferred name (e.g., *rabbit*) and a dispreferred name (e.g., *bunny*). Two trials before it was the participant’s turn to name a dual-name picture, the experimenter named the same picture, either using its dispreferred name (50% of trials) or its preferred name (50% of trials). We operationalized lexical alignment as the odds of a participant using a name that matched the name previously used by the experimenter and asked whether this differed between patients and controls. We also asked whether alignment was more effortful in patients, as indexed by naming response times.

Our second goal was to understand the relationship between lexical alignment in the patient groups and the two major constructs that have been linked to alignment — lexical priming and mentalizing/perspective-taking. To index *lexical priming* in language production, we embedded another subset of “single-name” trials within the picture naming game and measured the difference in the time it took for participants to name repeated pictures (named first by the experimenter) versus unrepeated (named first by the participant). To index *perspective-taking*, participants carried out a standardized test of mentalizing—Section 3 of The Awareness of Social Inference Test (TASIT; McDonald et al., 2003)—in which they watched videos of characters using lies and/or sarcasm and answered yes/no questions about these characters’ intentions. We selected this measure because it is commonly used to assess mentalizing in schizophrenia (e.g., Roberts & Penn, 2009; Sparks et al., 2010) , and is psychometrically sound (McDonald et al.,

2006; Pinkham et al., 2016), with direct relevance to how people make inferences about others' mental states in conversation. For each of these constructs, we first asked whether there were differences in performance between patients and controls. Based on the previous literature, we predicted that, while lexical production priming would be spared (or even increased) in patients, mentalizing would be impaired. We then asked whether, *within* the patient groups, individual differences in each measure predicted individual differences in lexical alignment. We predicted that, while impaired mentalizing would be linked to decreased lexical alignment, greater lexical priming would be linked to increased lexical alignment.

Methods

Participants

Sixty-four outpatients were recruited from McLean Hospital—32 with schizophrenia/schizoaffective disorder, and 32 with bipolar disorder (DSM-IV). Thirty-one control participants with no history of neuropsychiatric disorders, who were not taking psychoactive medication, were recruited by advertisement. All were native American English speakers with no (self-reported) history of learning or language disability. All provided informed consent following the guidelines of the Partners Institutional Review Board and were compensated for their time. Patients were assessed within eight weeks of the experimental session using several standardized scales (see Table 1 and Appendix A for further details about clinical characterization and recruitment).

	<i>Control</i>	<i>Schizophrenia</i>	<i>Bipolar</i>
N	31	32	32
Gender (M F NB) ^a	17 14 0	22 10 0	16 16 0
Race (AI A HPI AA W MR NR) ^b	2 0 0 4 23 1 1	0 0 0 7 23 2 0	0 1 0 3 28 0 0
Age	38.52 (13.8)	44.19 (11.1)	33.53 (9.8)
Parental SES ^c	49.96 (12.3)	42.88 (15.8)	53.82 (9.2)
Premorbid Verbal IQ ^d	110.21 (8.2)	100.58 (10.88)	109.04 (9.2)
TASIT ^e	55.16 (4.5)	46.5 (9.6)	52.9 (6.12)
Duration of Illness (years)	-	21.39 (11.4)	15.19 (11.2)
PANSS: General ^f	-	32.16 (7.5)	28.84 (6.0)
PANSS: Positive ^f	-	16.25 (5.6)	11.48 (3.8)
PANSS: Negative ^f	-	15.59 (5.4)	11.32 (3.8)
YMRS: Total ^g	-	11.31 (5.7)	8.87 (8.3)
Antipsychotics (T AT) ^h	-	8 21	7 16
CPZ Equivalent (mg) ⁱ	-	433.48 (340.8)	402.33 (301.1)

Table 1. Means are indicated, with standard deviations in parentheses. ^aM: Male, F: Female. ^bAI: American Indian/Alaska Native, A: Asian, AA : African American/Black, W: White, MR: More than one race, NR: Unknown or not reported. ^cHawaiian/Pacific Islander was also considered, but no participants identified with this group. ^dSES: Parental Socioeconomic Status, assessed using the Hollingshead Index (Hollingshead, 1965). ^eThe Awareness of Social Inference Test (McDonald et al., 2003). ^fPANSS: Positive and Negative Syndrome Scale (Kay et al., 1987). ^gYoung Mania Rating Scale (Young et al., 1978). ^hT: Typical, AT: Atypical. ⁱCPZ: Chlorpromazine equivalent, see Appendix A for details of calculation. The three groups were matched on gender, $X^2(2, 95)=2.484$, $p = .289$, but it was not possible to match all groups on age, parental socioeconomic status, or premorbid verbal IQ. To account for potential confounds, these three variables were covariates in all analyses in which Group was a predictor of interest.

The Collaborative Picture Naming Game

The game was set up as shown in Figure 1. On each trial, a picture of an object, animal, or person appeared on each screen. The two players alternated between naming the picture (a “production trial”) and listening to their partner name the picture, then stating whether their picture was the “same” or “different” (a “comprehension trial”). Below we describe the participants’ production trials used in our analyses. See Appendix A for a full description of the comprehension trials and development of stimuli.

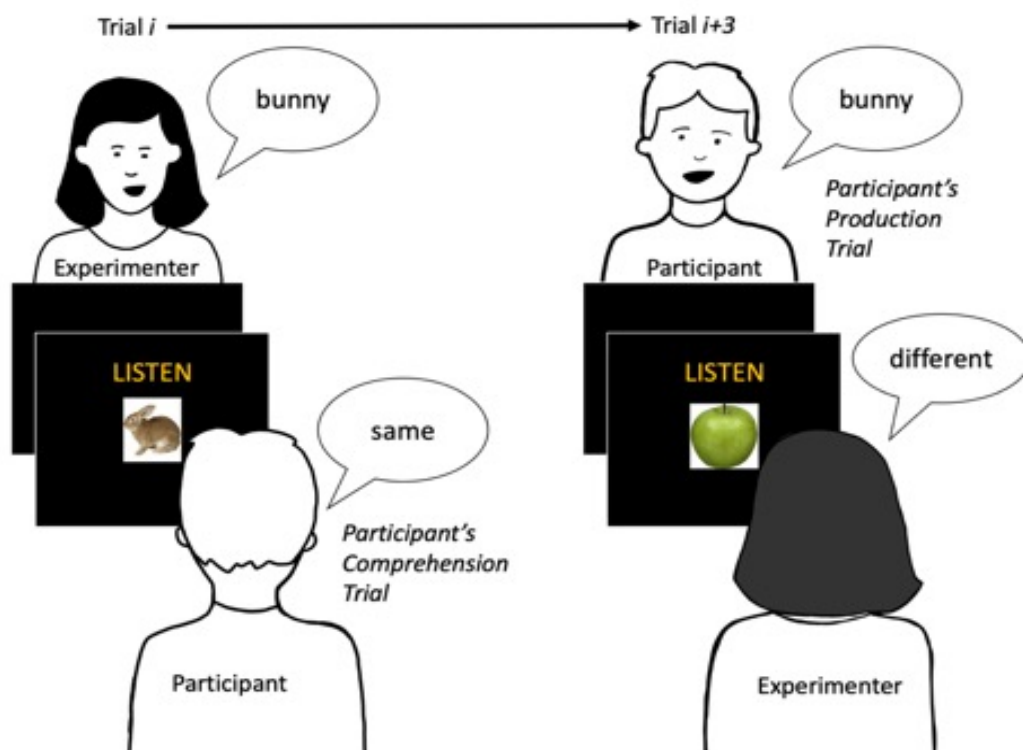


Figure 1. Setup of the cooperative picture naming game. The participant and experimenter faced one another, each looking at their own computer screen, which was not visible to the other player. On the i^{th} trial, the experimenter named a dual-name picture, and the participant, looking at his own screen, then stated whether the picture that he was looking at was the “same” or “different.” After two intervening trials (the $i+3^{\text{th}}$ trial), the participant named the same picture (here, using the same name the experimenter used), and the experimenter said “same” or “different.”

Of these 110 production trials, 20 were used to probe alignment. Each used a picture that had two possible names: a *preferred* name (e.g., “rabbit”), which was more frequent (SUBTL-US corpus; Brysbaert & New, 2009), and a *dispreferred* name (e.g., “bunny”) that was nonetheless judged to be highly acceptable (mean acceptability rating of 6.5 on a 1-7 scale). These dual-name stimuli were developed by carrying out two norming studies, described in Appendix A.

During the experiment itself, each of the 20 dual-name pictures was initially named by the experimenter and, two trials later, by the participant. In 10 trials, the experimenter used the picture’s dispreferred name; in the other 10 trials, the experimenter used the picture’s preferred name. These conditions (preferred, dispreferred) were counterbalanced across lists so that no participant named the same picture twice, but across participants, each dual-name picture was preceded half the time by its dispreferred name and half by its preferred name.

Twenty additional production trials were used to probe lexical production priming. In these trials, the pictures had a single name that was matched in frequency to the dispreferred name of one of the dual-name pictures. In 10 of these trials, the picture was repeated; that is, two trials back, the participant first heard its name spoken by the experimenter. In the other 10 trials, the picture was unrepeated: the participant named the picture first. Again, these conditions (repeated, unrepeated) were counterbalanced across lists. The remaining 70 production trials were filler trials with high frequency single-name pictures. All were named for the first time by the participant.

Before the experiment itself, participants received instructions stressing the cooperative aspect of the game and asking them to describe the pictures as specifically as possible, using a

single word (e.g., “rabbit” rather than “animal” or “small grey rabbit”). Participants then completed a practice phase. Trials were self-paced, and the experiment was split into five blocks. Between blocks, the percentage of trials that were correctly identified as “same” or “different” appeared on the screen, providing feedback on joint performance to encourage cooperation. The session was recorded, and the names and naming times were extracted following procedures described in Appendix A.

TASIT

After the picture naming game, participants completed Section 3 of TASIT (McDonald et al., 2003). The score was calculated as the number of correct responses (maximum: 68).

Statistics

All analyses were conducted using mixed-effects regression (see Appendix A for details). In all analyses, we included the maximal identifiable random effects structure, which generally consisted of random intercepts at the subject and item levels, and random slopes for all predictors of interest that varied within subjects and items, respectively. Thus, we accounted for by-item and by-subject variance, while controlling Type I error (Barr et al., 2013).

Results

Is lexical alignment spared or impaired in people with schizophrenia and bipolar disorder?

To assess the ability of participants to align to the experimenter, we created a binary variable, *Alignment*, which specified whether, for dual-name pictures, participants produced names that matched or mismatched the names produced by the experimenter. We then used

logistic regression, with Alignment as the dependent variable, to determine whether the log odds of alignment exceeded chance levels within each group. Given our design, the probability of match by chance was 0.5 (or equivalently, log odds = 0). In contrast, a participant who lexically aligned with the experimenter would be more likely to use the dispreferred name after hearing the experimenter use the dispreferred name, increasing the probability that the name they used matched that of the experimenter (log odds > 0). We found that, in all three groups, the log odds of producing a name that matched the name used by the experimenter were significantly greater than zero (Table 2).

To determine whether this alignment effect differed between the three groups, we carried out a between-group analysis with Alignment as the dependent variable and with Group (schizophrenia vs. bipolar vs. controls) as the between-subjects predictor of interest. Three additional subject-level variables were included as “control” regressors: Premorbid Verbal IQ (Blair & Spreen, 1989) , Age, and Parental SES (Hollingshead, 1965). We found no evidence of any between-groups differences in the log odds of alignment (Table 2, Fig. 2A). However, because a non-significant effect does not imply evidence *for* the null hypothesis, we then confirmed the absence of alignment differences between groups with non-inferiority testing (Lakens et al., 2018) and computed Bayes Factors, reported in Appendix A

Log Odds of Lexical Alignment – Within and Between Groups

	Predictor	Estimate (log odds)	SE	z	p	Sig.
<i>Within Controls</i>	Intercept	1.08	0.17	6.31	0.00	***
<i>Within Schizophrenia</i>	Intercept	1.01	0.13	7.58	0.00	***
<i>Within Bipolar</i>	Intercept	1.39	0.17	8.44	0.00	***
<i>Between Groups</i>	Intercept	1.02	0.17	6.12	0.00	***
	Group (Schiz.)	0.09	0.19	0.47	0.64	
	Group (Bipolar)	0.34	0.18	1.94	0.052	.

Table 2. ^Marginally significant with slightly greater alignment in the bipolar group. Logistic regression analyses examining the log odds of lexical alignment within and between groups. Effects of predictors of interest are shown. Note that alignment within each group is given by the Intercept of each model, which describes the log odds of the participant’s word choice matching that of the experimenter. See Appendix A for full results.

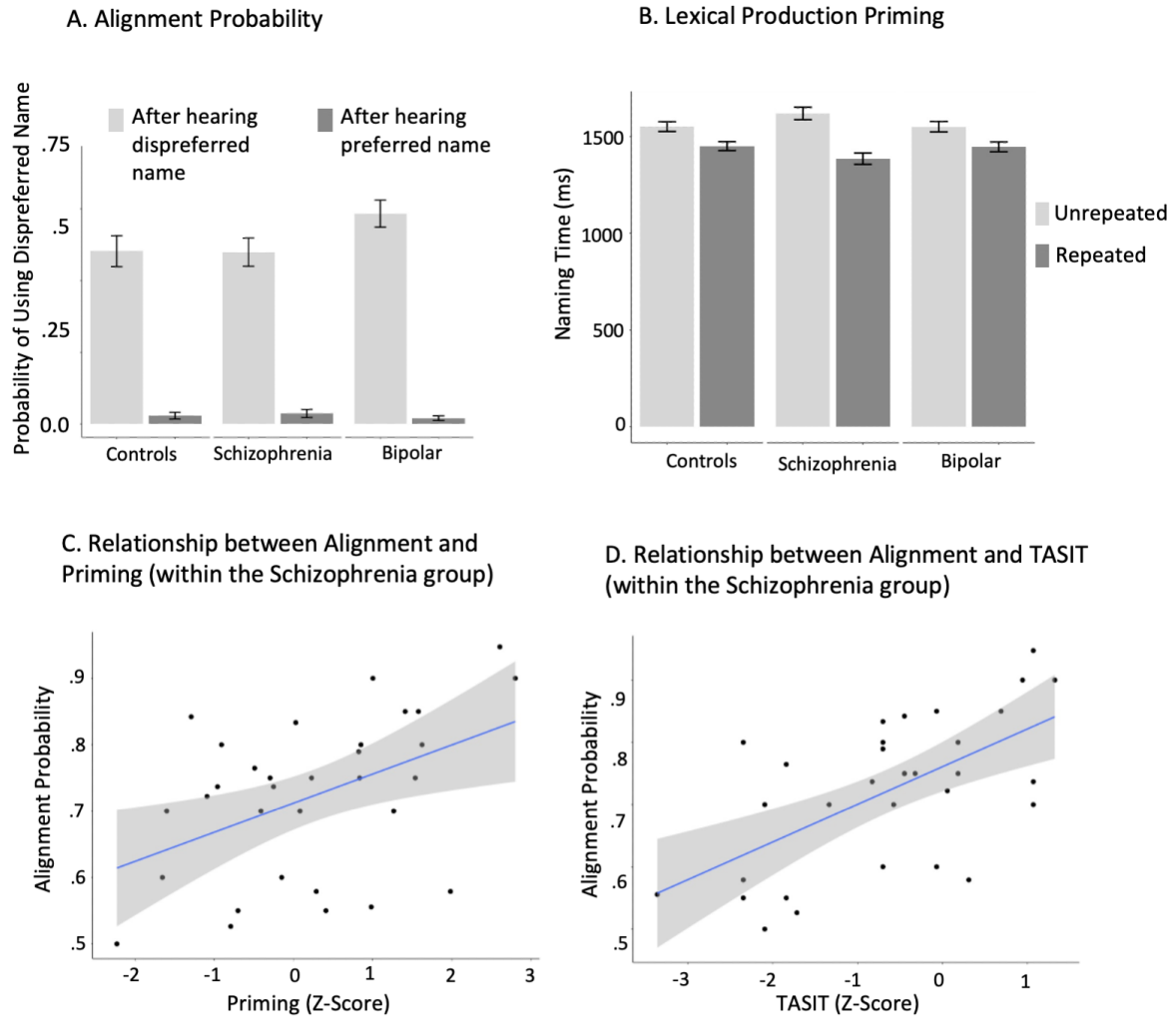


Figure 2. (A) Probability of using the dispreferred name after hearing the experimenter use the dispreferred name versus the preferred name. Error bars represent one standard error. (B) Lexical production priming: predicted naming times for repeated versus unrepeated low frequency single-name pictures, after controlling for demographic and lexical factors and for mean response time. Error bars represent one standard error. (C) Positive correlation between lexical production priming effects and alignment probabilities within the schizophrenia group. Points represent actual data, whereas the line represents the estimated regression coefficient (Est. = 0.23; SE = 0.09; $p=0.01$) and the shaded region around it represents the predicted standard error. (D) Positive correlation between mentalizing (TASIT scores) and alignment probabilities within the schizophrenia group. Points represent actual data, whereas the line represents the estimated regression coefficient (Est. = 0.32; SE = 0.08; $p=0.00$) and the shaded region around it represents the predicted standard error.

We then addressed the possibility that alignment was more effortful in the patient groups than in the control group. To do this, we extracted naming times for a subset of dual-name targets in which the experimenter had first produced the dispreferred name (e.g., “bunny”), and carried out linear mixed effects regression to compare participants’ times to name pictures on which they did align (e.g., produced “bunny”; 52% of trials) and did not align (e.g., produced “rabbit”; 48% of trials) to the experimenter. Log-transformed naming times served as the dependent variable, and we also included the average naming time for each participant as a subject-level predictor, to account for the fact that patients’ longer overall naming times can lead to spurious inflations of naming time differences (Chapman et al., 1994). We found no significant difference in naming times between aligned and non-aligned trials in any group, nor was the interaction between Group and Alignment significant (see Table 3).

Effect of Alignment on Naming Times of Dual-Name Trials						
	Predictor	Estimate	SE	t	p	Sig.
<i>Within Controls</i>	Alignment	-0.01	0.03	-0.20	0.84	
<i>Within Schizophrenia</i>	Alignment	0.00	0.04	0.03	0.97	
<i>Within Bipolar</i>	Alignment	-0.05	0.04	-1.23	0.24	
<i>Between Groups</i>	Alignment	0.02	0.04	-0.20	0.64	
	Group (Schiz.)	0.20	0.07	3.01	0.00	**
	Group (Bipolar)	0.12	0.06	2.04	0.05	.
	Alignment*Group (Schiz.)	0.02	0.07	2.52	0.81	
	Alignment*Group(Bipolar)	-0.07	0.05	-1.01	0.21	

Table 3. Effect of Alignment on log-transformed naming times within and between groups. Effects of predictors of interest are shown. See Appendix A for full results. Statistical significance was assessed using a type-III sums of squares estimation, with p-values estimated using the Satterthwaite approximation (Satterthwaite, 1946).

What are the relationships between lexical alignment, lexical production priming and mentalizing ability within the patient groups?

Our second goal was to understand how individual differences in lexical production priming and mentalizing ability might drive individual differences in lexical alignment. We began by asking whether these two constructs differed between groups.

Lexical Production Priming. We used mixed effects linear regression to determine whether there were differences in production priming between groups (again, log-transformed and with by-subject mean naming times as a covariate). We found a significant main effect of Repetition, with faster naming times to repeated than non-repeated single-name trials (see Table 4, Fig. 2B). We also found a Group by Repetition interaction, driven by a larger production priming effect in the schizophrenia group than the control group, which persisted above and beyond the variance accounted for by Premorbid Verbal IQ, Parental Socioeconomic Status, and Age. There was no difference in production priming between the bipolar group and the controls.

Effect of Repetition on Naming Times

	Predictor	Estimate	SE	t	p	Sig.
<i>Within Controls</i>	Repetition	-0.08	0.02	-3.75	0.00	**
<i>Within Schizophrenia</i>	Repetition	-0.13	0.03	-4.39	0.00	***
<i>Within Bipolar</i>	Repetition	-0.08	0.02	-4.31	0.00	***
<i>Between Groups</i>	Repetition	-0.07	0.03	-2.57	0.01	*
	Group (Schiz.)	0.04	0.02	1.53	0.13	
	Group (Bipolar)	0.00	0.02	0.14	0.89	
	Repetition*Group (Schiz.)	-0.07	0.03	-2.39	0.02	*
	Repetition*Group (Bipolar)	-0.01	0.03	-0.30	0.76	

Table 4. Effect of Repetition (repeated vs. non-repeated) on log-transformed naming times within and between groups. Effects of predictors of interest are shown. See Appendix A for full results. Statistical significance was assessed using a type-III sums of squares estimation, with p-values estimated using the Satterthwaite approximation (Satterthwaite, 1946).

TASIT. As expected, TASIT scores were significantly lower in the schizophrenia group (mean = 46.50, sd = 9.61) than in controls (mean = 55.16, sd = 4.50; Est. = -3.57, SE = 1.62, $p < 0.05$). The difference in scores between the bipolar group (mean = 52.90, sd = 6.12) and the controls was marginally significant (Est. = -2.59, SE = 1.48, $p = 0.08$).

Predictors of Alignment within Patient Groups. Above, we showed that both lexical production priming and mentalizing were abnormal in patients with schizophrenia. To test our hypothesis that both constructs would predict individual differences in lexical alignment amongst patients, we carried out regressions *within* the schizophrenia group. We found that better TASIT scores and larger lexical production priming effects both predicted a greater probability of alignment (Table 5, Fig. 2C, D). TASIT and lexical production priming were uncorrelated with one another ($r = -0.03$, $p = 0.19$), despite adequate variance in the distribution of these scores, and predictors remained significant when we included them in the same regression model, suggesting that they each accounted for unique sources of variance in alignment probability.

In the bipolar group, production priming again predicted alignment (with a larger priming effect predicting greater alignment), but the effect of TASIT on alignment was non-significant (Table 5). The effect of production priming remained significant when TASIT was included in the same model. Note that the distribution of TASIT scores in the bipolar group was heavily right-skewed, suggesting a ceiling effect may have accounted for the absence of an effect of TASIT on alignment.

For exploratory analyses of the relationship between alignment and symptoms, see Appendix A.

Effect of Mentalizing and Lexical Production Priming on Log Odds of Alignment

	Predictor	Estimate (log odds)	SE	z	p	Sig.
<i>Within the Schizophrenia Group</i>						
<i>Model 1: TASIT Only</i>	TASIT	0.32	0.08	4.00	0.00	***
<i>Model 2: Priming Only</i>	Priming	0.23	0.09	2.65	0.01	**
<i>Model 3: TASIT +</i>	TASIT	0.30	0.08	3.57	0.00	***
<i>Priming</i>	Priming	0.18	0.08	2.20	0.03	*
<i>Within the Bipolar Group</i>						
<i>Model 1: TASIT Only</i>	TASIT	-0.20	0.19	-1.04	0.30	
<i>Model 2: Priming Only</i>	Priming	0.42	0.15	2.85	0.00	**
<i>Model 3: TASIT +</i>	TASIT	-0.29	0.18	-1.63	0.10	
<i>Priming</i>	Priming	0.47	0.15	3.19	0.00	**

Table 5. Logistic regression analyses examining the effect of mentalizing (TASIT) and lexical production priming (Priming) on the log odds of alignment in patients. Effects of predictors of interest are shown. See Appendix A for full results.

Discussion

We used a collaborative picture-naming task to investigate lexical alignment in people with schizophrenia, bipolar disorder, and healthy controls. We found evidence of lexical alignment in all three groups, with no differences between the patient groups and the controls. In our analyses of response times, we also found no evidence that alignment to the dispreferred names was more effortful in either of the patient groups than the controls. Interestingly, however, the schizophrenia group showed significant differences from controls in two constructs that have been linked to lexical alignment: *reduced* mentalizing, operationalized by performance on the TASIT, and *increased* production priming, indexed as faster times to name repeated *versus* unrepeatd single-name pictures. Moreover, within the schizophrenia group, individual differences in both these constructs predicted variability in lexical alignment such that better TASIT scores and increased production priming predicted an increased probability of lexical

alignment.

Given that schizophrenia and bipolar disorder have both been linked to impairments in multiple aspects of social cognition and communication (e.g., Bora & Pantelis, 2016), the preservation of lexical alignment in both groups is striking. It is, however, consistent with a small body of work showing that other types of alignment—on “politeness” (Stewart et al., 2008) and syntactic structure (Dwyer et al., 2020)—are relatively spared in schizophrenia. It is also consistent with work showing that, despite clear social atypicalities, people with autism spectrum disorder exhibit typical alignment across multiple levels of linguistic representation (e.g., Allen et al., 2011; Slocombe et al., 2013; Nadig et al., 2015; Branigan et al., 2016). Taken together, these findings suggest that, rather than being globally impaired, people with these neuropsychiatric disorders have windows of spared social and communicative functioning.

Given the known link between lexical alignment and perspective-taking in healthy adults (Branigan et al., 2011), it is particularly interesting that lexical alignment in schizophrenia was spared, despite patients’ poor performance on the TASIT. We suggest two possible reasons for this dissociation. The first is that the TASIT tapped into a form of perspective-taking that was distinct from the more implicit “audience design” function that has been proposed to influence lexical alignment. In the TASIT, participants watch videos and make judgments about the intentions of the depicted characters who use non-literal language (e.g., sarcasm). Making explicit judgments requires offline reasoning and introspection that may not be engaged during everyday interaction (see Frith, 2004 for discussion).

On the other hand, the TASIT also taps into implicit top-down perspective-taking mechanisms that may be shared with lexical alignment. For example, selecting a dispreferred (non-literal) meaning over a preferred (literal) meaning requires the ability to infer another

person's communicative intent. Analogously, in the present alignment task, the *selection* of a dispreferred name (*bunny*) over a preferred name (*rabbit*) requires participants to represent the experimenter's mind and infer that alignment would facilitate communicative success (Branigan et al., 2011). The implicit online use of top-down pragmatic information is impaired in schizophrenia (Rabagliati et al., 2018), and such impairments have long been linked to communicative difficulties (e.g., Harrow et al., 1989; Pawelczyk et al., 2017). We therefore favor another explanation for why lexical alignment was spared in the schizophrenia group — that patients *compensated* for their impaired top-down mentalizing with enhanced bottom-up priming mechanisms.

In support of this account, the schizophrenia group showed a significantly larger lexical production priming effect than controls. This is consistent with some previous studies reporting enhanced lexical priming in schizophrenia during both comprehension (e.g., Spitzer et al., 1994; Kreher et al., 2008) and production (Kuperberg et al., 2018). In this previous work, however, “hyperpriming” was observed only under highly automatic conditions in which the prime preceded the target by less than 400ms within a single trial. While there are some studies showing that, when targets are separated from primes by several trials (analogous to the production lexical priming paradigm used here), repetition priming during comprehension is preserved (e.g., Clare et al., 1993; Sponheim, 2004; Doniger et al., 2001), the present findings provide the first evidence that, during language production, lexical repetition priming is enhanced in schizophrenia.

As noted in the Introduction, lexical alignment in healthy adults is thought to be mediated by both bottom-up priming (Pickering & Garrod, 2004) and top-down mentalizing (Branigan et al., 2011) mechanisms. For example, after hearing the experimenter produce a dispreferred name

(*bunny*), bottom-up influence may increase the level of activation of this relatively infrequent lexical representation so that it is equal to that of the more frequent lexical representation (*rabbit*). Therefore, when the participant names the picture, top-down perspective-taking may play an additional role in selecting the dispreferred name to achieve alignment, as described above. We suggest that, in schizophrenia, enhanced bottom-up priming raised the activation level of the repeated dispreferred lexical representation such that it was even higher than that of the preferred lexical representation, allowing patients to produce the dispreferred name and align with the experimenter, without relying on top-down mentalizing.

Further evidence that enhanced bottom-up priming might have compensated for reduced top-down mentalizing during lexical alignment in the patient groups comes from our examination of individual differences. In the schizophrenia group, we found that patients with worse mentalizing performance were less likely to align with the experimenter, while patients with larger-than-average lexical production priming effects showed greater alignment. We also showed that lexical production priming predicted lexical alignment within the bipolar group, although in this group, there was less variability in TASIT scores and no evidence of a relationship between TASIT and alignment.

Limitations and Open Questions

The interpretation offered above—that lexical alignment was preserved in schizophrenia because increased bottom-up priming compensated for reduced top-down mentalizing—leads to several testable predictions.

First, it predicts that lexical alignment may not *always* be spared in schizophrenia. More complex social situations are likely to impose greater demands on top-down mentalizing (Reitter

& Moore, 2014), calling on producers to dynamically up- or down-regulate alignment based on their beliefs about different partners (cf. Branigan et al., 2011). Using a non-social lexical priming paradigm, we recently showed that, in contrast to healthy controls, people with schizophrenia failed to implicitly adapt to changes in linguistic environments by engaging in top-down prediction (Sharpe et al., 2020). This raises the possibility that, even though patients' greater reliance on bottom-up priming allowed them to achieve similar *baseline* levels of alignment to controls, they may be less able to flexibly *adapt* their alignment in more dynamic social contexts (Nadig et al., 2015; Branigan et al., 2016). Thus, understanding how patients use top-down beliefs to facilitate alignment has implications for predicting psychosocial outcomes, as patients who are unable to flexibly adapt their speech to their partners' are likely to have difficulties in social communication.

A limitation of the present study is that we were unable to link heterogeneity of socio-cognitive impairments/mechanisms of lexical alignment with heterogeneity in specific symptom profiles amongst patients because our sample size was too small with too little variance in psychotic symptoms. Of particular interest is the relationship between lexical alignment and positive thought disorder, which is the most obvious clinical manifestation of linguistic communicative difficulties in schizophrenia (American Psychiatric Association, 2013). Some previous studies have linked thought disorder to impairments in taking the listeners' perspective into account (Rochester et al., 1977; Hoffman et al., 1982; Hoffman, 1986; see also Rutter, 1985; Harrow et al., 1989), and to more severe impairments in mentalizing ability (Frith and Corcoran, 1996). There is also preliminary evidence that thought-disordered patients exhibit impaired alignment at other levels of linguistic representation (syntax & description types; Dwyer et al., 2020). It will therefore be important for future studies using larger samples to investigate how

individual symptoms predict individual differences in lexical alignment.

Conclusions

To conclude, our results suggest that, at least in structured communicative environments like the present experiment, lexical alignment is unlikely to contribute to communication difficulties in chronic schizophrenia or bipolar disorder. We argue that these typical patterns of lexical alignment were supported by preserved—and in some cases increased—bottom-up mechanisms, which balanced out deficits in top-down perspective-taking. This interpretation raises important questions about how imbalances in bottom-up and top-down mechanisms of alignment may impact communication during more complex social interactions.

Chapter 4. Patterns of Lexical Predictability in Healthy Adults' Speech: Evidence from GPT-2³

Introduction

Naturalistic language production has historically been difficult to study quantitatively. Fields like pragmatics and conversational analysis have developed important ideas about how communication is guided and structured under real-world conditions, but less is known about the specific neurocognitive and computational underpinnings of language production. Yet, understanding language production is inarguably important, particularly for populations where language output may be atypical, as in schizophrenia and autism spectrum disorders. Indeed, as we discuss in Chapter 2, and in the next chapter of this thesis, it is striking that, in schizophrenia at least, it is still not clear how best to even *describe* such language atypicalities, let alone how to characterize their underlying mechanisms.

For years, quantifying language production atypicalities present in schizophrenia has relied on clinical rating scales, which are quite subjective and lack consistency in how constructs are defined (see, for example, the definition of “Loose associations” in the Thought, Language and Communication Scale (TLC; Andreasen, 1986) vs. in the Thought and Language Disorder Scale (TALD; Kircher et al., 2014). Recently, however, a set of computational techniques known as Natural Language Processing (NLP) have begun to drastically change the landscape of this literature. That is, a number of studies have begun using computational measures to quantify aspects of language production thought to be predictive of psychosis, in order to discriminate patients from controls and predict onset of psychosis (for review, see Chapter 2, and also,

³ This study was done in close collaboration with former undergraduate RA Santiago Noriega.

Corcoran et al., 2020).

As we discuss in Chapter 2, it is likely that these techniques can be similarly used to understand *why* language production goes wrong in psychosis as well - the mechanisms that govern language atypicalities. In particular, we suggest that a measure known as lexical predictability provides an ideal solution for beginning to understand these atypicalities. Lexical predictability, in contrast with the other computational measures that have been investigated in schizophrenia, captures a holistic relationship between a particular word and its context, accounting for constraints at all levels of linguistic representation, from syntactic constraints to more overarching message-level constraints. Lexical predictability has also, already, been established as one of the best predictors of processing difficulty in language comprehension (Smith & Levy, 2013; Staub, 2015; DeLong, Urbach, & Kutas, 2005; see Kuperberg & Jaeger, 2016, for discussion), and has been shown to relate linearly to behavioral processing time (Brothers & Kuperberg, 2021).

More specifically, however, we suggest that lexical predictability as a function of context *length* - the amount of context available with which to predict the target word - may be especially informative about impaired language production in schizophrenia, in part due to its face validity with Bleuler's concept of "loosening of associations" (Bleuler, 1950). Indeed, it has been suggested that deficits in using contextual information online - particularly information which must be maintained over longer timescales (i.e. more global information) - drive the disorganized speech observed in some patients (e.g. Spitzer et al., 1994; Spitzer, 1997; Kuperberg, McGuire, & David, 1998). Thus, one hypothesis would be that, relative to healthy controls, the relationship between lexical predictability and context length is atypical in patients, particularly when context lengths are longer.

However, there are two obstacles to testing such a hypothesis. The first of these obstacles is that estimating lexical predictability has historically been done using the Cloze task (Taylor, 1953) - a fill-in-the-blank task, which is completed by a sample of healthy adults (usually around 50), whose responses are then used to approximate a probability distribution over possible continuations for a given context (everything that comes before the blanked-out “target” word). Cloze is very time- and labor-intensive and therefore ill-suited to computing lexical predictability, word by word, for a large sample of speech data (and large samples are generally needed for patient populations, whose data tends to be quite noisy). Now, though, there are large language models like OpenAI’s GPT-2 (Radford et al., 2019), which can compute lexical predictability quickly and easily. The second obstacle is that the idea of “atypical” use of global context in patients implicitly assumes that healthy adults use local and global context in consistent, predictable ways to produce language - an assumption that seems reasonable but has historically been difficult to test directly.

In the present study, we address both of the obstacles described above by leveraging the large language model GPT-2 in order to compute lexical predictability for each word in three separate corpora. We first use a corpus of natural-like vignettes to compare how different measures computed from GPT-2 compare with those computed from a human Cloze task, with an eye toward assessing GPT-2 as a tool for studying lexical predictability as a function of context length, and identifying possible areas of weakness for the model. We also use two corpora of transcribed natural speech to test whether, and how, lexical predictability relates to context length, with the prediction that, in line with pressures to make speech more predictable for one’s listener, lexical predictability should increase significantly, and relatively monotonically, with increasing context length.

Methods

Corpus 1: Semi-Naturalistic Vignettes

We first compared GPT-2 probability measures with those derived from a human Cloze task (Taylor, 1953), in a corpus made up of 179 vignettes of four sentences each, which were designed as stimuli for another study (Alexander, in prep). These stories were designed to resemble naturalistic discourse, while probing different degrees of predictability - thus making this corpus an ideal test set for comparing GPT-2 probability measures with human intuitions. Each sentence had between eight to fourteen words, and all the sentences related to the others within the same set. For example:

“The boy scouts were excited about their hiking trip. They were going to spend an entire week in the wilderness honing their skills. They had been on similar adventures before and knew just what to do. After arriving in new territory, the scouts would always explore the area.”

Calculation of GPT-2 Cloze Measures. GPT-2 (Radford et al., 2019) is a powerful large language model developed by OpenAI. In part, the novelty of GPT-2 as a predictor of human language comes from its architecture, which employs a technique known as ‘transformers’. The model is trained on a large set of text: WebText, a diverse library of over 8 million human-curated online documents of 40GB in total.

In this study, we used the GPT2-Large model from HuggingFace’s Transformers repository (Wolf et al., 2020), which has 36 layers, 345 million parameters, and a vocabulary of 50,257 tokens. We used the GPT-2-for-Psycholinguistic-Applications repository

(github.com/samern92/GPT-2-for-Psycholinguistic-Applications) to extract probability distributions over possible continuations for a variety of contexts, in two different corpora, as described below.

For each word in each vignette, given all the available context to the left of the target word, GPT-2 probability distributions were extracted to compute four measures commonly used to describe constructs related to predictability in language: Lexical Predictability (the probability of a word occurring given its prior linguistic context), Lexical Surprisal (the negative log-transform of lexical predictability), Constraint (the probability of the most likely continuation given the context), and Entropy (the weighted mean surprisal value over possible continuations to the given context; intuitively, a measure of “uncertainty” about the upcoming continuation). Because computing entropy over the entire 50,257-token vocabulary of GPT-2 was computationally infeasible, and also unlikely to be comparable with measures of entropy obtained from Cloze data (which is inherently limited to N continuations, where N is the number of subjects who participated in the Cloze task), we instead computed entropy over the 50 most likely continuations, normalizing these probabilities such that their sum equaled 1.

Human Cloze Estimates. To obtain analogous measures based on human intuitions about language, we additionally performed a Cloze task, again for each word in each vignette. Using Amazon Mechanical Turk, we presented subjects with one word of the vignette at a time, asking them to then fill in the following word. Once they responded, the correct word would then appear (indicating to subjects whether their guess was correct or incorrect). They would be then be able to enter the next word, and so on. Subjects were instructed to only enter a single word at each time point. (Refer also to the Cloze Sentence Completion Task described by Schnoebelen

and Kuperman, 2010) for details of this word-by-word Cloze paradigm).

Vignettes were divided up into nine non-overlapping lists. Individual subjects were allowed to complete more than one list but were not allowed to complete any particular list more than once. Eligibility for the study required that subjects: had an IP address within the United States, had completed 100 or more HITs on Amazon Mechanical Turk, had an overall HIT acceptance rate of greater than 98%, and were native English speakers (self-reported). This resulted in the collection of 477 datasets (mean per list = 52.56). Workers gave written informed consent prior to participating in the study, in accordance with the procedures of the Institutional Review Board of Tufts University.

Of these 477 datasets, 4 were excluded for being less than 50% complete. Another 96 datasets were excluded for poor accuracy (<10%). This threshold was determined by examining the distribution of accuracy scores across participants, which was strongly bimodal; manual examination of the datasets with less than 10% accuracy revealed that few of these responses appeared to be made in good faith (rather, low accuracy tended to result from participants entering nonsense responses, e.g. random strings of letters or the same word over and over). An additional 6 datasets were excluded for unusually high accuracy (accuracy > 50%), which suggested that participants had done the experiment more than once. Amongst the remaining 361 datasets (185 female, 168 male; mean age = 33.54, SD = 8.76) mean accuracy was 25.15% (SD = 4.31%).

Each (word-by-word) response was screened for spelling errors using the Python library PyEnchant; responses that were flagged as misspelled were checked manually. When it was clear what the misspelled word was supposed to be (e.g. “suprise” for “surprise”), the response was corrected. If it was unclear, the response was left as is. Responses which were not word-like (e.g.

strings of random letters) were excluded from the data. After cleaning, each word trial had an average of 40.06 individual responses ($SD = 3.91$, $min = 28$, $max = 46$). The probability of a given word, in a given context, was taken to be the number of respondents who entered that word, divided by the total number of included responses. These data were used to calculate lexical predictability (the proportion of responses that were correct), surprisal, constraint, and entropy for each word. We note that because the proportion of correct responses was in some cases zero, due to the finite number of discrete responses collected, we “smoothed” probabilities for the purpose of computing surprisal by adding 1 to both the total number of responses and the number of correct responses.

Questions and Analysis. We tested the relationship between each pair of analogous GPT-2 and Cloze-derived measures using Pearson correlations in *R* (R Core Team, 2022). We report correlation coefficients, t-values, and 95% confidence intervals for these analyses. When obvious non-linearity was detected in plots, we additionally computed Spearman correlations. For each pair of measures, we also performed simple regression analyses to compute a line of best fit in order to better quantify the degree to which GPT-2 was over- or under-estimating a measure relative to the Cloze data. Note that, due to the presence of a number of outliers for GPT-2 Surprisal, we limited the data included in these analyses to datapoints where GPT-2 Surprisal was within 3 standard deviations of the mean.

Corpus 2: Naturalistic Speech from the BNC

We next examined the relationship between predictability and context. Because the corpus used above was designed to probe a range of predictability values, and therefore was not

necessarily reflective of a typical relationship between context and predictability, we instead turned to a large corpus of transcribed natural speech: the British National Corpus (BNC; BNC Consortium, 2007), an extensive and diverse collection of written and spoken language in British English. Here, we used, specifically, the spoken section of this corpus, in order to examine patterns of lexical predictability in language output produced “online”.

The spoken section of the BNC is subdivided into two sub-corpora: the DEM-BNC and the CG-BNC. In the present study, we analyze these sets of texts separately, as they index somewhat different types of language production. The DEM-BNC is a “demographically-sampled” collection of everyday, informal conversations among volunteers of different demographics. The CG-BNC, in contrast, is a “context-governed” collection consisting of more formal conversations or speeches generally centered around a specific topic. For example, a casual conversation between friends would be representative of DEM-BNC transcript, whereas subjects covered in the CG-BNC range from soccer punditry to history lectures.

The DEM-BNC makes up about 40% of the Spoken corpus, and the CG-BNC the remaining 60%. In total, thousands of speakers and 4.2 million words are transcribed in the Spoken BNC. The BNC also includes XML markers corresponding to the speaker saying each word. This allowed us to define a “Speech Segment” as any contiguous (un-interrupted) segment of speech spoken by a single speaker. GPT-2 measures were always computed based on only the context within each individual Segment, i.e. never including language output from a different speaker.

For each the first 500 words of each Speech Segment, we used GPT-2 to calculate Lexical Predictability, using two different context levels. First, we computed predictability based on all the available context within the speech segment. However, because, in this measure,

context length is inherently confounded by the word's position in the segment, we additionally computed sentence-level predictability, which was based on only the prior context within the present sentence (as defined by end-of-sentence punctuation given in the transcripts).

For each word in each Speech Segment, we coded several additional variables to help tease apart the relationship between predictability and different levels of context. This included Word Position, which indexed the position of each word within the speech segment (1st, 2nd, and so on) and, for Full Context measures, was equivalent to the length of the context in words, minus 1; Word Position in Sentence, which indexed the position of each word within the containing sentence (thus, for Sentence Context measures, this was equivalent to the length of the context in words, minus 1); Sentence Position in Segment, which indexed the position of each sentence within the segment (1st sentence, 2nd sentence, and so on); and, finally, Segment Length, which indexed the length of the Speech Segment in words.

Analysis. To test our primary questions, which are described in the next section, *R* (R Core Team, 2022) was used to perform all analysis. For linear mixed effects regression (LMERs) analyses, we used packages *lmerTest*, version 3.1-3 (Kuznetsova, Brockhoff, & Christensen, 2015) and *lme4*, version 1.1-33 (Bates et al., 2015). For all LMERs, we used the maximal random effects structure justified by the data per the recommendations of Barr et al. (2013), which generally meant modeling random intercepts by subject and item (speech segment), and random slopes with respect to each for all predictors of interest which varied within subject and item, respectively (see Results for model specifications).

Model convergence was generally not an issue but when it did occur, our first step was to change the optimizing algorithm employed by *lme4*. If the model still did not converge, we

simplified the model by setting the correlation parameters for random effects to zero (indicated by the double bar “||” syntax in the model specifications). All models converged after these steps.

In our analyses, predictors indexing context length (Word Position and Word Position in Sentence) were log-transformed to increase linearity. Additionally, to normalize residuals, correct for some heteroskedasticity, and make our results more interpretable, we used the negative log-transform of lexical predictability (thus converting to the information-theoretic measure “Surprisal”)

Results

Corpus 1: Semi-Naturalistic Vignettes

As described in the Methods section, we obtained GPT-2 and Cloze-based predictability measures for each word in a corpus of semi-naturalistic vignettes, with the goal of assessing whether GPT-2 measures can approximate the patterns in word probability distributions obtained from crowd-sourced Cloze data.

We first compared the raw lexical probabilities output by GPT-2 with the probabilities derived from the Cloze task, i.e. “Lexical Predictability”. We found that, as expected, there was a very strong positive correlation between Cloze Lexical Predictability and GPT-2 Lexical Probability (Pearson’s $r = 0.80$, $t(7018) = 112.83$, 95% CI [.79, .81]), with GPT-2 generally underestimating probabilities relative to Cloze (Figure 3). This underestimation appeared more severe for words which were more predictable.

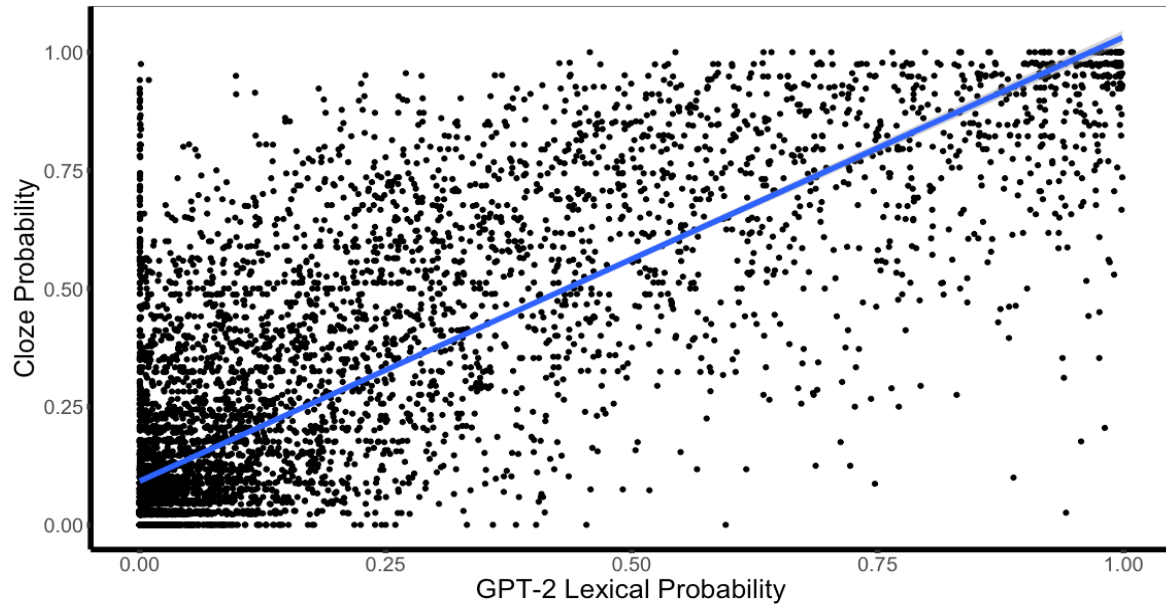


Figure 3. Blue line represents line of best fit based on OLS regression with Cloze Lexical Predictability as the dependent variable. Intercept: .09 (SE = .002, $t = 35.75$, $p < .001$); Slope: .94 (SE = .008, $t = 112.83$, $p < .001$).

In the subset of Surprisal data (see Methods), the correlation between Cloze Surprisal and GPT-2 Surprisal was, similarly, moderately strong and positive, though with a somewhat smaller correlation coefficient than for the lexical probabilities (Pearson's $r = .68$, $t(6928) = 78.144$, 95% CI [.67, .70]). Inspection of the scatter plot (Figure 4) suggested this may have been due to a far greater range of surprisal values at low probabilities (the granularity of which, in Cloze, was inherently limited by the number of participants).

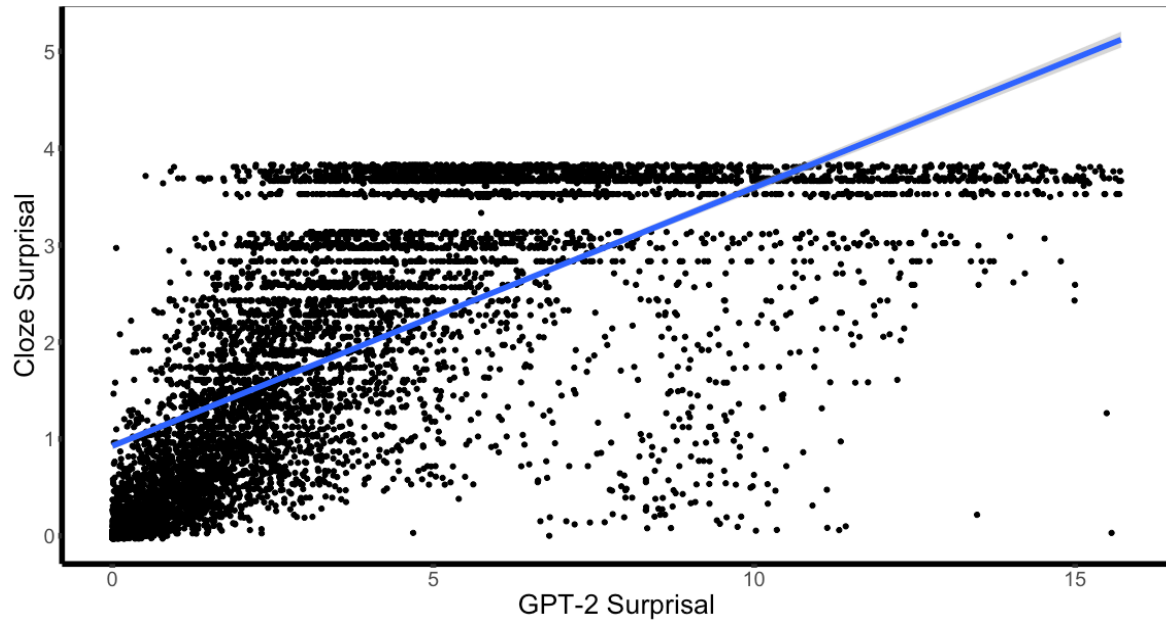


Figure 4. Blue line represents line of best fit based on OLS regression with Cloze Surprisal as the dependent variable. Intercept: .95 (SE = .018, $t = 51.94$, $p < .001$); Slope: .27 (SE = .003, $t = 78.16$, $p < .001$).

GPT-2 Constraint was also positively correlated with Cloze Constraint (Pearson's $r = .65$, $t(7018) = 71.56$, 95% CI [.64, .66]). In line with the patterns observed for lexical predictability, GPT-2 generally underestimated constraint relative to the Cloze data (Figure 5).

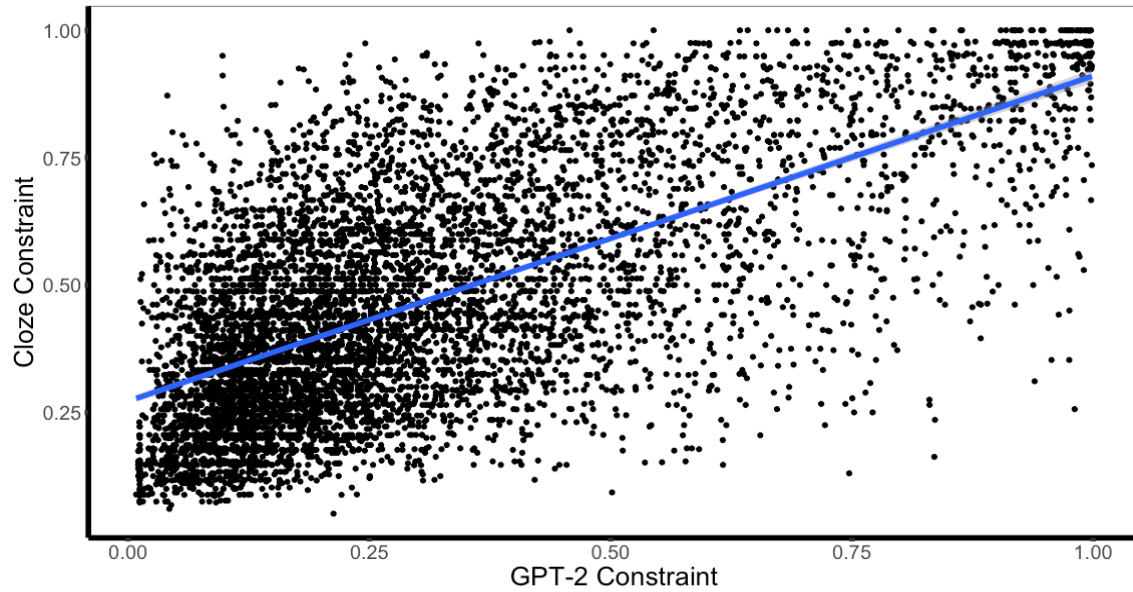


Figure 5. Blue line represents line of best fit based on OLS regression with Cloze Constraint as the dependent variable. Intercept: .27 (SE = .003, $t = 78.94$, $p < .001$); Slope: .64 (SE = .009, $t = 71.56$, $p < .001$).

Finally, GPT-2 Entropy was, also, positively correlated with Cloze Entropy (Pearson's $r = .73$, $t(7018) = 88.51$, 95% CI [.72, .74]). Interestingly, visualization of this relationship (Figure 6) showed that, while the measures were particularly well-correlated when entropy was low, there was also some non-linearity and non-monotonicity at higher levels of entropy. Due to this non-linearity, we additionally computed Spearman's rho for this pair of variables, finding $\rho = .49$, $S = 22.85$. This suggests that GPT-2 entropy measures may be less appropriate to use when entropy is expected to be relatively high. It is also worth noting that GPT-2-based Entropy is consistently overestimated compared to that obtained from the Cloze task; however, whether this had more to do with GPT-2's tendency to underestimate lexical probability, or the inherent limitations of computing this measure from Cloze data, is unclear.

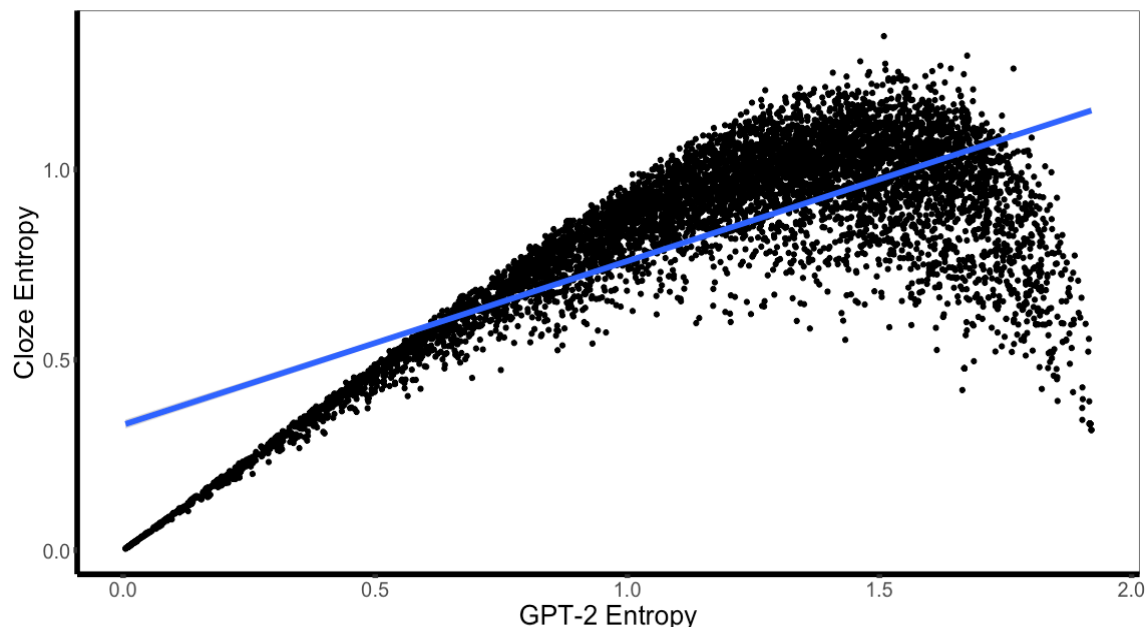


Figure 6. Blue line represents line of best fit based on OLS regression with Cloze Entropy as the dependent variable. Intercept: .33 (SE = .006, $t = 54.58$, $p < .001$); Slope: .43 (SE = .004, $t = 88.91$, $p < .001$) (Though, we note that the results of this regression are likely unreliable due to the strong non-linearity of the relationship.)

Corpus 2: Naturalistic Speech from the BNC

In this corpus, we first asked whether there were reliable patterns in surprisal with increasing discourse-level context length, expecting that predictability should increase (and surprisal decrease) with increasing context length (I.e. Word Position). We additionally modeled speech segment length as a covariate.

Predictability Based on Full Context. Based on visualization of the data, we observed that in both the DEM-BNC and the CG-BNC, on average, lexical predictability appeared to increase with Word Position (Figure 7). Moreover, plotting predictability against the natural logarithm of Word Position seemed to mostly linearize the relationship (Figure 8).

In both corpora, linear mixed effects regression analysis confirmed a significant main

effect of Log Word Position on Surprisal (Table 6) such that, when utterance length was held constant at the mean, for each unit of increase in Log Word Position, Surprisal decreased by approximately .71 (CG; equivalent to an approximately 103% *increase* in Lexical Predictability) or .78 (DEM; equivalent to an approximately 118% increase in Lexical Predictability).

In the CG corpus, Segment Length had a positive effect on Surprisal, such that, at the mean of Log Word Position, Surprisal was overall lower for longer segments of speech; this effect was also positive, but only marginally significant, in the DEM corpus. However, there was, in both corpora, a significant, positive interaction between Segment Length and Log Word Position on Surprisal, such that the slope of the decrease in surprisal with respect to Log Word Position was flatter (less negative) for longer Segment Lengths.

Model: Surprisal (full context) ~ Log Word Position*Segment Length + (1 + Log Word Position | Speaker_ID) + (Log Word Position | Segment)

Effect of Word Position and Speech Segment Length on Full Context Surprisal

CG-BNC	Estimate	SE	t	p	Sig.
Intercept	3.86	0.01	304.94	0.00	***
Log Word Position	-0.72	0.01	-77.36	0.00	***
Speech Segment Length	0.18	0.02	10.28	0.00	***
Log Word Position*Speech Segment Length	0.22	0.01	16.21	0.00	***
DEM-BNC	Estimate	SE	t	p	Sig.
Intercept	4.63	0.02	219.13	0.00	***
Log Word Position	-0.78	0.02	-40.68	0.00	***
Speech Segment Length	0.07	0.04	1.99	0.08	.
Log Word Position*Speech Segment Length	0.07	0.02	4.14	0.00	***

Table 6. Predictors of interest shown in bold.

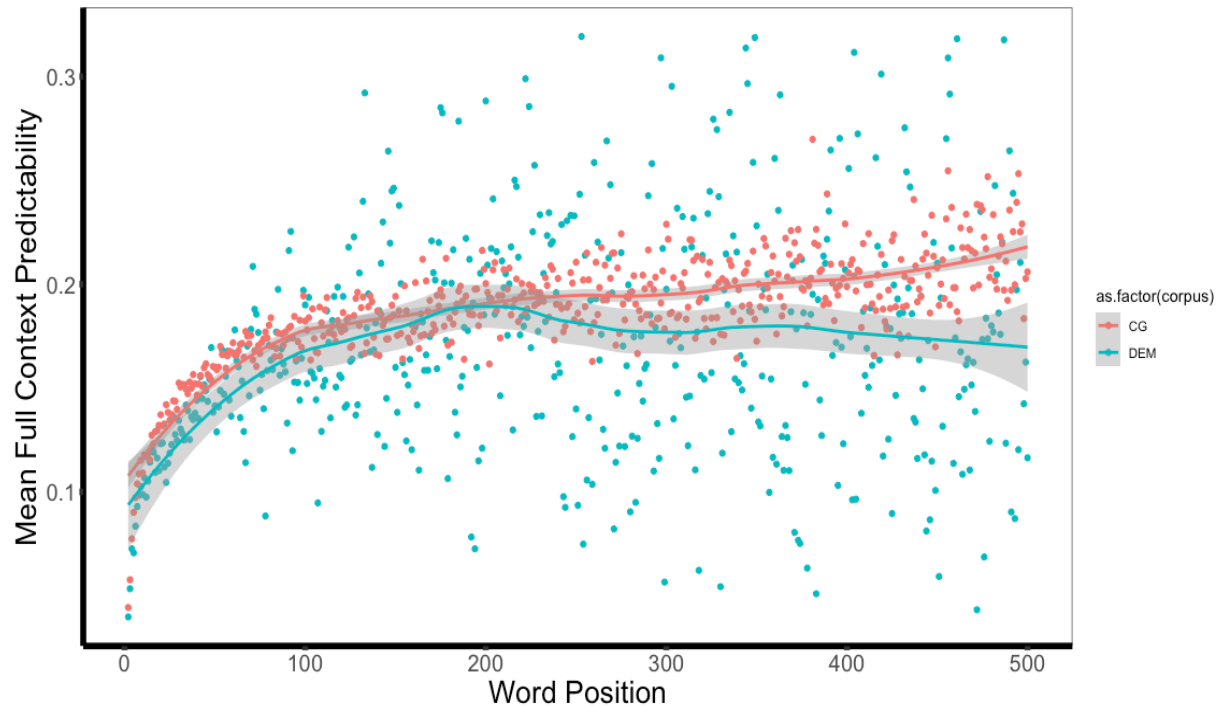


Figure 7: Mean Full Context Predictability plotted against Word Position. Lines represent smoothed data via local regression; shaded regions represent 95% CI.

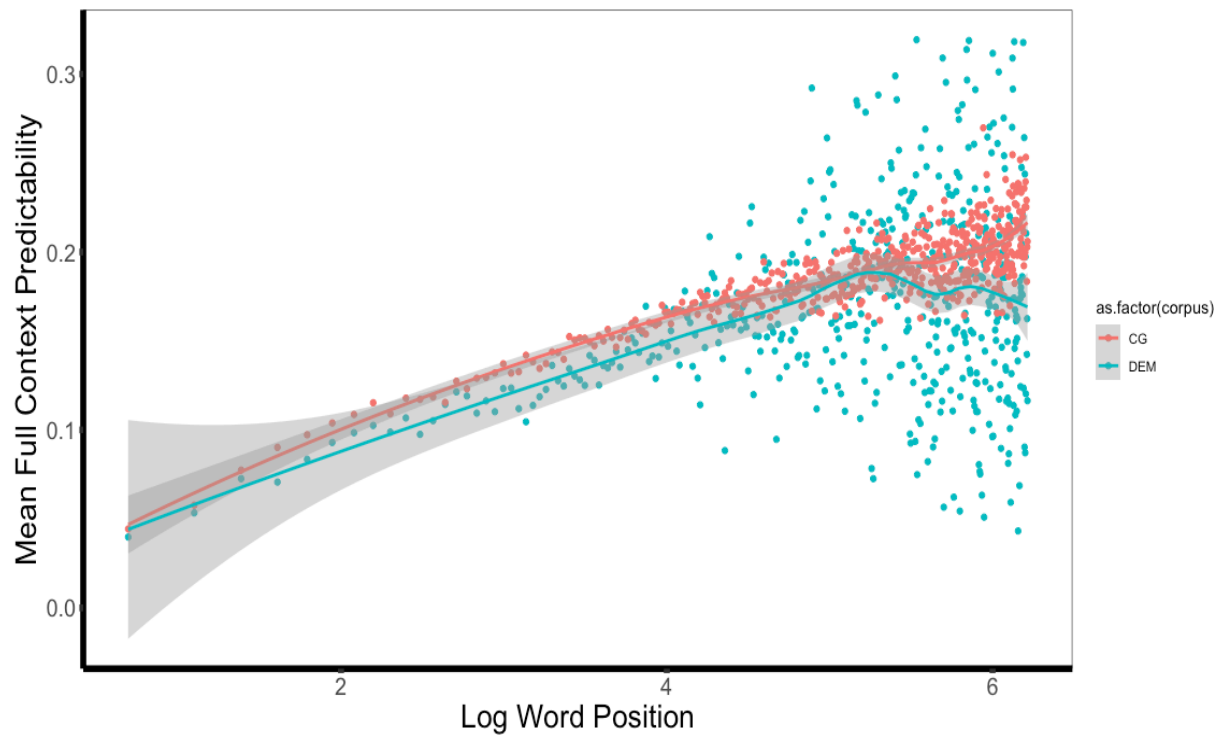


Figure 8. Mean Full Context Predictability plotted against Log Word Position, Lines represent smoothed data via local regression; shaded regions represent 95% CI.

Predictability Based on Sentence Context. We next ran a similar analysis looking at the effects of context within sentences, in order to distinguish effects of word position in discourse from effects of the length of the context *available* to the model. Here Word Position in Sentence served as the proxy for context length, and we again expected to see a negative effect of this variable on surprisal. Sentence Position was additionally modeled as a covariate.

Visualization suggested a strong positive effect of Word Position in Sentence on predictability within both corpora (Figure 9). The analysis confirmed, within both corpora, a large negative effect of Word Position in Sentence on surprisal (Table 7), such that, independent of the position of the word within the subject's speech segment, varying the context length available to the model significantly modulated surprisal. In both corpora, the main effect of Sentence Position in Segment, and the interaction between Sentence Position and Word Position in Sentence were non-significant.

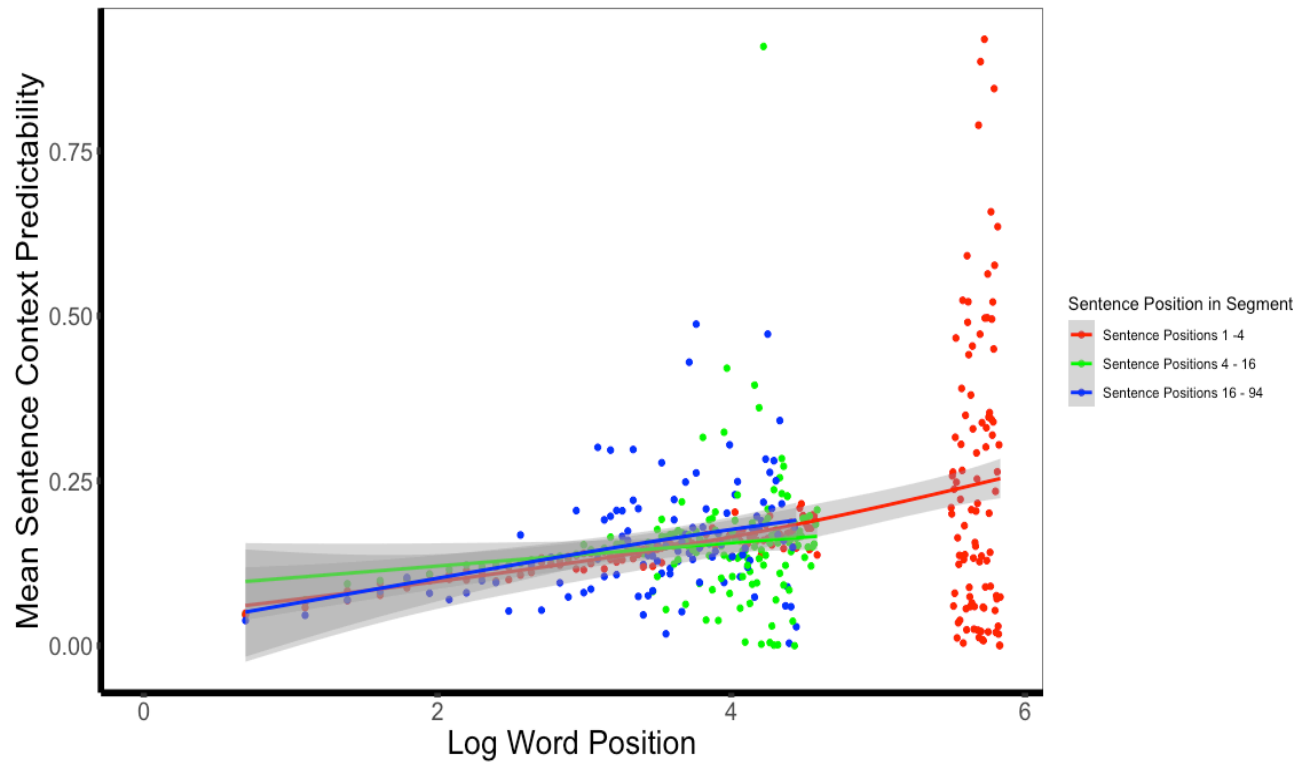


Figure 9. Mean Sentence Context Predictability plotted against Log Word Position, Lines represent smoothed data via local regression; shaded regions represent 95% CI.

Model: Surprisal (sentence context) ~ Log Word Position in Sentence*Sentence Position in Segment (1 + Log Word Position in Sentence | Speaker_ID) + (1 + Log Word Position in Sentence | Segment)

Effect of Word Position in Sentence and Sentence Position in Segment on Sentence Context Surprisal					
CG-BNC	Estimate	SE	t	p	Sig.
Intercept	4.86	0.01	372.77	0.00	***
Log Word Position in Sentence	-0.64	0.01	-69.00	0.00	***
Sentence Position in Segment	-0.02	0.01	-1.55	0.12	
Log Word Position in Sentence*Sentence Position in Segment	0.01	0.01	1.22	0.22	
DEM-BNC	Estimate	SE	t	p	Sig.
Intercept	5.21	0.02	267.21	0.00	***
Log Word Position in Sentence	-0.61	0.02	-34.20	0.00	***
Sentence Position in Segment	0.00	0.02	-0.07	0.95	
Log Word Position in Sentence*Sentence Position in Segment	0.01	0.02	0.25	0.81	

Table 7. Predictors of interest shown in bold.

Discussion

In the present study, we first show that measures extracted from the large language model GPT-2 are tightly correlated with analogous measures computed based on human judgments from a Cloze task. This was particularly true for measures of lexical predictability, which achieved the highest correlation coefficient. Such a finding is promising for studies which aim to answer questions about lexical predictability as a function of context in schizophrenia (as we do in Chapter 5).

Cloze, a type of fill-in-the-blank task, has long been considered the gold standard for computing measures related to lexical probabilities. Indeed, until the advent of transformers-based models like GPT, there was no language model that could even approach the power of human judgments for estimating lexical predictability from discourse contexts. Here, however,

we show that measures of lexical predictability, surprisal, constraint, and entropy derived from GPT-2 are all strongly, positively correlated with the analogous measures computed from a Cloze task. Among these, lexical predictability stands out as particularly accurate, with a correlation coefficient of over .8. Though surprisal, a direct transformation of lexical predictability, was not as highly correlated, this may have been due to a lack of granularity in the Cloze probabilities, and a higher correlation may be expected given more responses to the Cloze task.

Together, these findings suggest that, with enough data, GPT-2 should be more than adequate for approximating lexical probability measures in natural speech. Our data is in line with previous assessments of GPT-2, which have shown that it guesses blanked-out words with high accuracy (Radford et al., 2019). Since the time of writing, several other studies have reported similar correlations between GPT and Cloze measures: Hao and colleagues (2020) report a correlation of .67 between GPT-2 lexical predictability and human-generated lexical predictability, based on random segments of language drawn from a larger text; Michaelov and colleagues (2023) report a correlation of -.53 between GPT-3 surprisal and Cloze probability; Szwedczyk & Federmeier (2022) report a correlation of .72 between GPT-2 probabilities and Cloze probabilities for word-final continuations in set of naturalistic stimuli; Goldstein and colleagues (2022) report correlations ranging from .46 to .79 between GPT-2 and Cloze probabilities; and Eisape, Zaslavsky, & Levy (2020) report a Kendall's Tau of .57 between GPT-2 and Cloze surprisal (in contrast to a correlation of .2 between Cloze surprisal and surprisal estimated from 5-gram probabilities). Additionally, a number of studies have found GPT-2 and GPT-3 surprisal to be strongly predictive of various behavioral (Goodkind & Bicknell, 2018; Wilcox et al., 2020; Boyce et al., 2023) and neural (Heilbron et al., 2019; Michaelov et al., 2023) measures which are known to be tightly linked to lexical predictability.

Our findings highlight, also, some discrepancies between GPT-2 measures and Cloze measures which may need to be taken into account when GPT-2 is used. For example, it is notable that correlation coefficients for Constraint and Entropy were lower than those for Lexical Predictability and Surprisal. This is, perhaps, not especially surprising, given that GPT-2 allocates probability mass over a vocabulary of non-word tokens, rather than words (Radford et al., 2019); thus, this already constitutes a major difference between GPT-2 output and that of a Cloze task. While there are likely ways to manipulate GPT-2’s output to get more accurate measurements, it is worth noting that “out of the box” probability distributions will not map precisely onto those approximated from human judgments. Additionally, it should be noted that GPT-2 is largely trained on web-based text, which means that it occasionally predicts elements—like an html tag, for example—that, obviously, would never occur in natural human speech output. The fact that GPT allocates probability mass to these elements may make its estimates of any measures which depend on the full probability *distribution* of continuations for a given context poorer. Given these caveats though, the fact that GPT-2 measures of Constraint and Entropy still achieve relatively strong positive correlations with their Cloze counterparts, particularly when uncertainty is lower, is quite impressive.

Interestingly, these discrepancies, though they appear to be largely the product of “deficiencies” in GPT-2, also raise the question of to what degree GPT-2 and human Cloze participants are really performing the same “task”, and whether GPT-2 might in some cases even be *more* appropriate than Cloze for approximating predictability-related measures. That is, GPT-2’s task is to generate a probability distribution over possible continuations, which then allows it to produce a continuation, or multiple continuations, to the text it has just seen. Humans are asked simply to select one specific word that best follows the context, but it is unclear how they

do this. Though some have theorized that humans perform Cloze tasks by implicitly generating probability distributions which they then sample from (Smith & Levy, 2011), there is evidence that Cloze performance is better described in terms of a “race” model (Staub et al., 2015). This latter interpretation means that the traditional logic behind the Cloze task - that aggregating information about the “best” completion from a large sample of humans should approximate the probability distribution underlying the possible continuations - may not always hold. Indeed, this is the argument of Michaelov and colleagues (2023), who argue that measures of surprisal extracted from large language models in fact are *better* predictors of neural activity than those extracted from human judgments. However, more research will be needed to confirm the degree to which this is truly the case, and in what contexts large language models may or may not “outperform” Cloze.

Given the success of GPT-2 in approximating human judgments about predictability, we next sought to understand the relationship between context length (the amount of context available to the model, or to the listener) and predictability/surprisal. We showed that lexical predictability/surprisal—varied in predictable ways within the speech of healthy adults. Specifically, lexical predictability increases (and surprisal decreases) reliably and monotonically as context length increases more and more.

We found, firstly, evidence from both the DEM-BNC and CG-BNC, that lexical predictability increases logarithmically with context length. This suggests that, as expected, healthy adults’ speech is guided by context across a wide temporal window. At the same time, more distant context is clearly less important than context which is closer to the target word, such that additional context results in “diminishing returns” for predictability. This is not surprising, as such a phenomenon has been predicted by models of “contextual relevance” which posit that

more distant contextual cues “lose” their relevance for predicting/producing upcoming linguistic signals (Qian & Jaeger, 2010). Our analysis of sentence-level predictability, in which the amount of context available to the model again significantly decreased surprisal, suggests that additional contextual information decreases surprisal regardless of the position of the word in the speaker’s speech. These patterns were, importantly, highly consistent across corpora.

Finally, we found that the increase in discourse-level predictability was modulated by the length of the discourse segment, with slower changes for longer segments of speech, highlighting the importance of accounting for this variable when looking at effects of context length on predictability.

Thus, large language models exhibit clear potential for helping us understand language production in healthy adults, and in patient populations. However, in order to best make use of these models, it is imperative that we first understand how their measures pattern in healthy, naturalistic speech, and how reliable those patterns are. Critically, we show that there are consistent patterns in predictability within natural healthy speech - normative “baselines” against which to test speech produced by patient populations.

These findings, therefore, have important implications for how we study language production in schizophrenia. Indeed, it has long been thought that the language impairments, both in comprehension and in production, observed in schizophrenia may derive from atypical sensitivity to context, particularly global context (Cohen & Servan-Schreiber, 1992; Spitzer, 1994). However, it has not been clear, (a) what, specifically, one would expect “atypical” use of context to look like, and (b) *how* one might be able to test such a hypothesis. In the present study, we show that, in healthy speech, speakers consistently make use of linguistic context across a wide temporal window, with predictability increasingly more or less monotonically as the

amount of available context increases.

One prediction, then, in line with the idea of a selective deficit in using *global* context, would be that for patients, use of context within a *shorter* temporal window (more local context) may be relatively intact; however, patients may be unable to appropriately use context from the larger window which governs healthy controls' speech. This would be in line with studies in language comprehension in schizophrenia, which have shown that patients can generally use minimal contexts as well as controls—whether a single word, as in semantic priming (for review, see Pomarol-Clotet, Oh, & Laws, 2018) or a clause or phrase (Swaab et al., 2013) - but struggle to use longer discourse contexts (Ditman & Kuperberg, 2007; Swaab et al., 2013; Boudewyn et al., 2017). In light of the findings in the present study, we would expect that predictability would still increase, with increasing context, in patient speech, but to a lesser degree than in healthy controls, and, importantly, that group differences in predictability given more global context would be disproportionately larger than group differences in predictability given more local contexts. In the next chapter, we will use GPT-2 estimates of lexical predictability to test this hypothesis.

Chapter 5. Lexical predictability in schizophrenia: Evidence for selective deficits in use of global context during language production

The study of language provides unique insights into higher-order cognition. This idea has long been recognized by psychopathologists. Indeed, the term “positive thought disorder” was coined to describe the highly disorganized speech produced by some patients with schizophrenia. Thought disorder has been linked to significant impairments in social functioning (Bambini et al., 2016; Marggraf, Lysaker, Salyers, & Minor, 2020), occupational functioning (Racenstein, Penn, Harrow, & Schleser, 1999), and overall quality of life (Tan, Thomas, & Rossell, 2014; Bambini et al., 2016). However, even in those patients without noticeable thought disorder, subtler impairments in language which may plausibly impair real-world communication have been observed. For these reasons, understanding how disorganized speech arises in schizophrenia is of great clinical importance. Despite this obvious importance, it remains unclear even how best to objectively quantify disorganized speech, let alone how to tease apart its cognitive underpinnings.

One hypothesis which may speak to both of these open questions harkens back to the earliest days of schizophrenia research when Bleuler proposed a “loosening of associations” as a fundamental feature of the disorder (Bleuler, 1950). He had observed, specifically, that some patients’ speech was dominated by associative links between individual words, at the expense of overall coherence. That is, patients may at times choose words that are related on a local scale even though they may seem globally inappropriate. Now, years later, imbalances in use of local versus global context have been documented across multiple perceptual and cognitive domains, including visuospatial processing (Hemsley, 2005) and attention (Bellgrove, Vance, & Bradshaw,

2003). Within language, this theory has received some support from a series of carefully controlled experiments in language comprehension. Such experiments have shown that, under relatively automatic conditions, patients can use single-word context (as in semantic priming; see Minzenberg, Ober, & Vinogradov, 2002, for review) within a very local temporal window to facilitate processing of upcoming words to the same extent that controls do. However, use of more global linguistic contexts which must be maintained over longer temporal windows (e.g. sentences) is relatively impaired (Ditman & Kuperberg, 2007; Swaab et al., 2013; Boudewyn et al., 2017).

So far, however, it has not been feasible to quantify use of local and global context in language production in people with schizophrenia. Instead, researchers have remained reliant on subjective ratings via various clinical ratings scales (e.g., Andreasen, 1986; Kircher et al., 2014). To develop a more principled understanding of whether, and how, disorganized speech is linked to use of context during speech production, one would need a measure of the relationship between individual words and the entire linguistic context that precedes them. Notably, there were some attempts to do this in the 1970s and 1980s using a technique called Cloze (Taylor, 1950) to derive measures of lexical probability based on human judgments (Salzinger, Portnoy, and Feldman, 1964; Salzinger, Pisoni, Portnoy, & Feldman, 1970; Manschreck, Maher, Rucklos & White, 1979; Ragin & Oltmanns, 1983; Ragin & Oltmanns, 1987). However, Cloze is a labor-intensive process, and therefore sample sizes were quite small, estimates of probability very imprecise (due to few participants in the Cloze task), and the context examined was generally bilateral, rather than limited to the preceding context. Most importantly, these studies did not test the critical hypothesis that use of global context is *specifically* impaired relative to use of local context (though see Salzinger and colleagues' 1970 study of bilateral global and local context in

a sample of 10 patients).

However, recently advents in computational power and machine learning have opened up myriad possibilities for describing and quantifying various aspects of language production. Indeed, over the last decade, natural language processing (NLP) methods have been extremely successful in discriminating patients from healthy controls (e.g., Elvevag et al., 2007; Mota et al., 2012, 2017; Mitchell et al., 2015; Corcoran et al., 2018; Iter & Jurafsky, 2018; Voleti et al., 2019) and predicting psychosis onset (e.g., Bedi et al., 2015; Corcoran et al., 2018; Spencer et al., 2021). Specifically, patients' speech has been shown to differ from that of controls on a number of measures, including semantic similarity, graph-theoretic measures of speech "connectedness", and NLP measures of referential cohesion (see Chapter 2, as well as Corcoran, 2020, for an excellent overview of this NLP work in schizophrenia). Moreover, some measures have been shown to correlate with specific symptoms; for example, atypical graph-theoretic measures are predictive of negative symptoms (Mota, Copelli, & Ribeiro, 2017). For the most part, these studies have taken a primarily descriptive approach - to classify or predict. There has been far less emphasis on using these types of measures to test more theory-driven hypotheses. In the present study, however, we present a novel approach which begins to unravel the cognitive underpinnings of disorganized speech. In a large sample of treatment-naïve first-episode psychosis patients and demographically matched healthy controls, we test the hypothesis that use of global linguistic context is selectively impaired in schizophrenia. To do this, we leverage a powerful, state-of-the-art large language model: GPT-3.

GPT-3 (Brown et al., 2020) is one of a series of large language models in the Generative Pretrained Transformer (GPT) series developed by OpenAI. GPT models have excelled at a wide variety of language-related tasks (Radford et al., 2019; Brown et al., 2020). Like many large

language models, it is trained to predict upcoming words on prior context. However, GPT is distinguished from other types of language models in part by its self-attention mechanisms, which allow it to more easily handle larger sequences of tokens (Vaswani et al., 2017). The ability to handle these larger sequences means that GPT is exceptionally good at taking into account longer, more complex linguistic contexts (Brown et al., 2020). Moreover, although the GPT series of models was not designed with biological plausibility in mind (see Millet et al., 2022 for discussion), GPT-3 output has been found to predict behavioral (Goodkind & Bicknell, 2018; Wilcox et al., 2020; Boyce et al., 2023) and neural (Heilbron et al., 2019; Michaelov et al., 2023) markers of language processing, suggesting that, much like humans, it is sensitive to constraints imposed by different levels of linguistic representation, across varying temporal windows (see also Cauchetaux, Gramfort, & King, 2023).

Most importantly for our purposes, GPT-3 allows the user to extract estimates of the probability of any individual word given its prior context in a measure known as lexical predictability. This enabled us to probe the relationship between each individual word in speech and the words produced previously. Based on our observations in Chapter 4, we know that lexical predictability should increase with context length. However, because here we are interested in how predictability changes *proportionally* with its magnitude, we here employed the information-theoretic measure lexical *surprisal*, the negative log of predictability. We first tested whether group differences could be detected in lexical surprisal given any and all available context, which would suggest that patients use context atypically. We then systematically manipulated the length of the context the model had access to, thus extracting measures of surprisal given a wide range of temporal windows, from very global (taking into account all, or a large portion of, the context that was produced) to very local (taking into account a single word,

or a few words, of the context that was produced). In this way, we tested the hypothesis that use of global context during language production is *selectively* impaired in schizophrenia. Finally, we tested the degree to which disproportionate impairments in use of global context were predictive of, and specific to, positive thought disorder.

Methods

Data Collection

Participants. We used a sample of speech data from a previously collected dataset (see Silva et al., 2021; Mackinley et al., 2021). These data were collected from 106 English-speaking participants: 36 Healthy Controls (HCs) and 70 treatment-naïve patients experiencing their first episode of psychosis (FEPs). All subjects were screened for the following exclusion criteria: meeting DSM-5 criteria for a substance-use disorder in the past year; a history of a major head injury (leading to a significant period of unconsciousness or seizures); significant, uncontrolled medical illness; intellectual/developmental disorders; greater than 2 weeks of lifetime antipsychotic exposure; and the inability to provide informed consent. Eleven subjects (10 patients; 1 control) were excluded prior to analysis due to missing Parental Socioeconomic Status (SES) information (see below)⁴. Demographic and clinical data for the final sample are provided in Table 8.

Patients were recruited from referrals to the PEPP (Prevention and Early Intervention for Psychosis Program) at London Health Sciences Centre, London, Ontario, Canada between April 2017 and July 2019. Patients were in the acute phase of illness and recruited upon referral, irrespective of hospitalization status, before antipsychotic treatment was established. All

⁴ Including these subjects in the analysis, with no control for Parental SES, did not qualitatively change the results.

participants received a consensus diagnosis from a minimum of 3 psychiatrists (2 research psychiatrists and the primary treatment provider from the PEPP clinic) after approximately 6 months on the basis of the best estimate procedure (as described in Leckman et al, 1982, and the Structured Clinical Interview for DSM-5, First, 2014). The 6-month consensus diagnosis confirmed that 92% ($n = 55$) of patients in the final sample met criteria for a schizophrenia spectrum disorder; the remaining 8% ($n = 5$) met criteria for affective psychosis⁵. Healthy controls were recruited from the community. None had a personal history of mental illness, nor any family history of psychotic disorders.

All participants gave informed written consent, following procedures approved by the Western University Health Sciences Research Ethics Board, London, Ontario, Canada.

⁵ Though these 5 subjects were included in the analyses reported below, we also conducted a sensitivity analysis to determine the effect of excluding them, and found that the results did not differ qualitatively: no effects of interest that were significant in the main analysis became non-significant when these subjects were excluded, nor did any non-significant effects of interest become significant.

Demographics and Clinical Measures

	Controls	Patients	
N	35	60	
Age	21.68 (3.22)	22.26 (4.34)	$t(87.56) = -0.74; p = .46$
SES	3.09 (1.36)	3.50 (1.27)	$t(67.36) = -1.47; p = .16$
Sex ^a	F:12; M: 23 ; NB: 0	F: 13; M: 47; NB: 0	$X^2 (1) = 1.22; p = .27$
PANSS-8 Total	8.00 (0.00)	25.79 (7.08)	$t(56) = -18.97; p < .05$
PANSS-8 Positive	3.00 (0.00)	12.48 (3.14)	$t(57) = -7.6795; p < .05$
PANSS-8 Negative	3.00 (0.00)	7.48 (4.45)	$t(57) = -22.95; p < .05$
TLI Total	0.31 (0.40)	1.6 (1.43)	$t(74.15) = -6.50; p < .05$
TLI Impoverishment	0.14 (0.25)	0.58 (0.72)	$t(80.62) = -4.20; p < .05$
TLI Disorganization	0.17 (0.26)	1.02 (1.26)	$t(67.97) = -5.07; p < .05$
DSST Score ^b	69.20 (11.00)	50.20 (11.00)	$t(82.04) = 7.48; p < .05$
Category Fluency ^c	24.70 (7.01)	18.20 (5.12)	$t(51.99) = 4.41; p < .05$
TMT Part B Score ^d	55.10 (14.90)	96.40 (74.40)	$t(45.40) = -3.50; p < .05$
Segment Length ^e	99.41 (35.23)	67.47 (31.63)	$t(65.17) = 4.42; p < .05$
Percent Content Words ^f	48.23% (3.56)	48.86% (4.82)	$t(87.71) = -0.73; p = .46$

Table 8. Standard deviations in parentheses; significant differences between groups shown in bold ^aF = Female; M = Male; NB = Non-Binary/Intersex; ^bDigit-Symbol Substitution Test, mean of Written and Oral scores; ^cCategory/Semantic Fluency: number of animal exemplars produced; ^dTrail-Making Test, Part B; ^eMean length of speech segment (in words); ^fPercent of words produced, per utterance, that were content words (vs. function words or disfluencies)

The Thought and Language Index. Speech samples were collected during administration of the Thought and Language Index (TLI; Liddle et al., 2002), an interview-based instrument designed to assess formal thought disorder under standardized conditions. All participants were asked to describe three pictures from the Thematic Apperception Test, for 1 minute each, after hearing specific instructions: “I am going to show you some pictures, one at a time. When I put each picture in front of you, I want you to describe the picture to me, as fully as you can. Tell me what you see in the picture, describe what you see in this image, and what you think might be happening.”

Responses were recorded, transcribed, and scored on eight domains which were used to compute two domain scores: (1) Impoverishment, which included poverty of speech, weakening of goal, and perseveration of ideas and (2) Disorganization in Thinking, which included

looseness, peculiar use of words, peculiar sentences, peculiar logic, and distractibility. The interviews and ratings were completed by trained graduate-level research assistants who were blinded to patient status. Portions of speech which were unintelligible were marked for later exclusion.

Symptom Assessment. In addition to assessment of thought disorder via the TLI, symptoms were assessed more generally using the Positive and Negative Syndrome Scale - 8 Item (PANSS-8; Lin et al., 2018), a condensed version of the interview-based PANSS (Kay, Fiszbein, & Opler, 1987) for psychosis. The PANSS-8 was carried out by two research psychiatrists who achieved a high degree of interrater reliability.

Cognitive Assessment. Cognition was assessed using three tasks: the Digit-Symbol Substitution Test (DSST), which is thought to index working memory and processing speed (Wechsler, 1981); Category Fluency (number of exemplars), which is thought to index access to semantic memory (Neill, Gurvich, & Rossell, 2014); and Part B of the Trail-Making Test (TMT), which is thought to index executive function (Arbuthnott & Frank, 2000). In the DSST, participants were instructed to match symbols to numbers based on a given key. Both oral and written portions were included, and scores averaged. In the Category Fluency task, participants were instructed to name as many exemplars of a particular semantic category (e.g. “animals”) as possible, and the number of exemplars was tallied. In the Trail-Making Test Part B, participants trace a line through consecutive numbers and letters, alternating between numbers and letters (1-A-2-B-3-C-and so on). The time taken to complete this connect-the-dots-style pattern correctly is recorded.

Speech Data Processing

Transcript Processing. We first segmented the picture description transcripts into contiguous segments of participant speech (the experimenter’s speech was not used in this study), which we will refer to as ‘discourse segments’, or simply ‘segments’. Individual utterances were then segmented into words by breaking the utterance prior to each space character. All punctuation except for apostrophes, hyphens, and sentence-final punctuation was removed.

Computing Lexical Predictability. To probe the lexical predictability of each word in the patient’s speech, we used the state-of-the art language model GPT-3, a transformers-based model developed by OpenAI (Brown et al., 2020). GPT-3 is trained to predict upcoming words based on a given context. Importantly, the predictions the model makes are only influenced by the context to the left of the word, restricting the model from accessing future contextual information. This restriction allows the model to form predictions in a way somewhat similar to how humans predict upcoming speech (given the incremental nature of spoken language). GPT-3 uses a token system which does not always align with word or morphological boundaries (e.g. the word “grasshopper” is tokenized by GPT-3 as “grass”, “hop”, and “per”). Thus, for words which were composed of multiple tokens, we obtained the lexical predictability for the whole word using the rules of conditional probabilities.

To obtain GPT-3 probabilities for each word in each discourse segment, we used the OpenAI API to interface with the “text-davinci-002” model (the second largest of the “davinci” set of models, which are currently the largest available pre-trained GPT-3 models). Importantly,

GPT-3 allows us to restrict the context available to the model, such that we are able to compute lexical predictability for a given word given a range of context lengths, from very narrow (a single preceding word) to very broad (many preceding words, up to $n - 1$ for the n th word in an utterance). For each word token we (a) extracted predictability given any and all context available up to that point (which we refer to as Overall Context) and then (b) extracted predictability while modulating context, providing the model with contexts ranging in length from 1 prior word to 50 prior words, truncating the available context when necessary. Thus Context Length ranged from 1 to 50 words, but was log-transformed prior to analysis in order to meet the assumption of linearity imposed by the GLM.

Data Cleaning

All predictability values were converted to surprisal by taking the negative log of the probability value. This was done for two reasons: first, the log-transformed values, compared to the original linear values, reduced heteroskedasticity and produced greater normality of the residuals, as assessed by visual inspection of diagnostic plots, and second, conducting our analysis in log space allowed us to assess group differences in predictability proportional to the magnitude of the probability (e.g., a 2% increase in probability) versus in linear space, where group differences are absolute, e.g. a .002 increase given each additional unit of context).

We then excluded any words that were marked as unclear in the transcription *and* all datapoints following the unclear word (as it was unknown how GPT-3 would deal with these annotations). We also excluded disfluencies such as “um”, “uh”, and “hmm”, as well as function words, as defined by NLTK’s list of “stopwords” (Bird, Klein, & Loper, 2009), as surprisal was not expected to vary much with context for these types of words, for either group. After these

exclusions were applied, a total of 19,162 word tokens (45.23% of the data) remained (Patients: 11473 tokens, 43.94%; Controls: 7689 tokens, 47.30%).

Finally, we eliminated outliers for surprisal, within each level of context. That is, words with surprisal outside 3 standard deviations from the mean surprisal at that level of context (e.g. the mean surprisal given 50 words of context, or the mean surprisal given Overall Context) were excluded. A mean of 67.93 (SD = 13.31) datapoints were excluded from each predictability variable.

Statistical Analysis

We used a series of mixed effects linear regressions to test our primary hypotheses, which are detailed in the Results. Regression analyses were conducted in R (R Core Team, 2020) using lmerTest, version 3.1-3 (Kuznetsova, Brockhoff, & Christensen, 2015) and lme4, version 1.1-33 (Bates et al., 2015) packages. All continuous predictors were z-scored.

Though the difference in Parental SES between groups was non-significant, because of strong *a priori* predictions about how this variable might affect surprisal (e.g., as a mediator of education level and verbal IQ), we included it as a covariate in all analyses. We also modeled Segment Length (in words) as a covariate in all analyses, based on the possibility that subjects may distribute information differently across discourse segments of different length (see, e.g., Chapter 4). Following the recommendations of Becker (2005) and Kalnins (2018), we followed up each analysis by repeating the same analysis without nuisance covariates. All significant effects on predictors of interest remained significant after removing covariates, and all non-significant effects of interest remained non-significant, unless otherwise noted.

In all analyses, we modeled the maximal random effects structure justified by the data

and our theoretical assumptions. This structure consisted of subject- and, when appropriate, word-level (meaning word token - a particular instance of a word occurring in a discourse segment) intercepts, and random slopes with respect to our predictors of interest when applicable (Barr, Levy, Scheepers, & Tily, 2013). In cases of model non-convergence, if changing the optimizing algorithm did not help, the random effects structure was simplified, beginning with removing correlation parameters for the most complex random effects (see model structures in Results for details). All models converged after either changing the optimization algorithm or removing correlation parameters.

Results

Question 1: Are people with schizophrenia impaired in using context to modulate surprisal of the words that they produce during language production?

As noted in the Introduction, healthy adults modulate the predictability (or surprisal) of the words that they produce based on prior context. In this case, this context constitutes the content of the picture they are describing and the full set of words that they have already produced to describe it. To test the broad hypothesis that people with schizophrenia are relatively impaired in their use of context to modulate lexical surprisal during language production, we used GPT-3 to extract two estimates of the surprisal of each word in each speech segment (a) WithContext surprisal — an estimate of surprisal that was based on any and all context (prior words) had previously produced within sample up until that point and (b) NoContext surprisal — an estimate of surprisal in which the context given to GPT-3 was a random sample of words (the same number of words as in the context that was actually produced) from an AI-generated description of an unrelated TAT picture (see Appendix B for details). This measure captured

aspects of the word in isolation (e.g. frequency) as well as any general content related to picture description more generally. We expected that, in both groups, WithContext surprisal would be smaller than NoContext surprisal. The critical question was whether patients atypicalities in their use of context to modulate surprisal of words produced (i.e. a smaller difference between NoContext and WithContext surprisal).

To test this hypothesis, we ran a linear mixed effects model with surprisal as the outcome measure. Predictors of interest were as follows: a between-subjects fixed effect of Group (Controls, Patients) in which Controls served as the reference level, a within-items fixed effect of Context Type (NoContext, WithContext) in which NoContext served as the reference level, and the interaction between Group and Context. As expected, we found a strong negative main effect of Context, such that surprisal was lower in the WithContext condition (see Table 9 for statistics). That is, controls were using the context they had previously produced to generate more predictable upcoming words. Critically, and confirming our hypothesis, the interaction between Context and Group was significant, and positive, suggesting that patient group failed to modulate the surprisal of their produced words based on prior context to the degree as healthy adults. Of note, the main effect of Group, which, in this analysis, indexed the difference between patients and controls in the NoContext Condition (the “reference” level), was numerically small and non-significant. This suggests that, in isolation of any context, the individual words patients produced were not inherently more surprising than those produced by controls (e.g. because they were less frequent or because they failed to engage in the task).

*Model Structure: Surprisal ~ Group*Context Type + SES* Context Type + Segment Length* Context Type + (1 + Context Type | Subject) + (0 + Context Type | Word)*

Effect of Group and Context Type (NoContext vs. RelatedContext) on Surprisal

	Estimate	SE	t	p	Sig.
Intercept	8.82	0.17	52.55	0.00	***
Group	0.01	0.09	0.15	0.88	
Context_Type	-4.38	0.21	-21.04	0.00	***
SES	0.03	0.04	0.76	0.45	
Segment Length	0.00	0.00	0.76	0.45	
Context_Type*Group	0.32	0.10	3.37	0.00	**
Context_Type*SES	0.02	0.05	0.37	0.71	
Context_Type*Segment_Length	0.00	0.00	-2.35	0.02	*

Table 9. Predictors of interest are shown in bold.

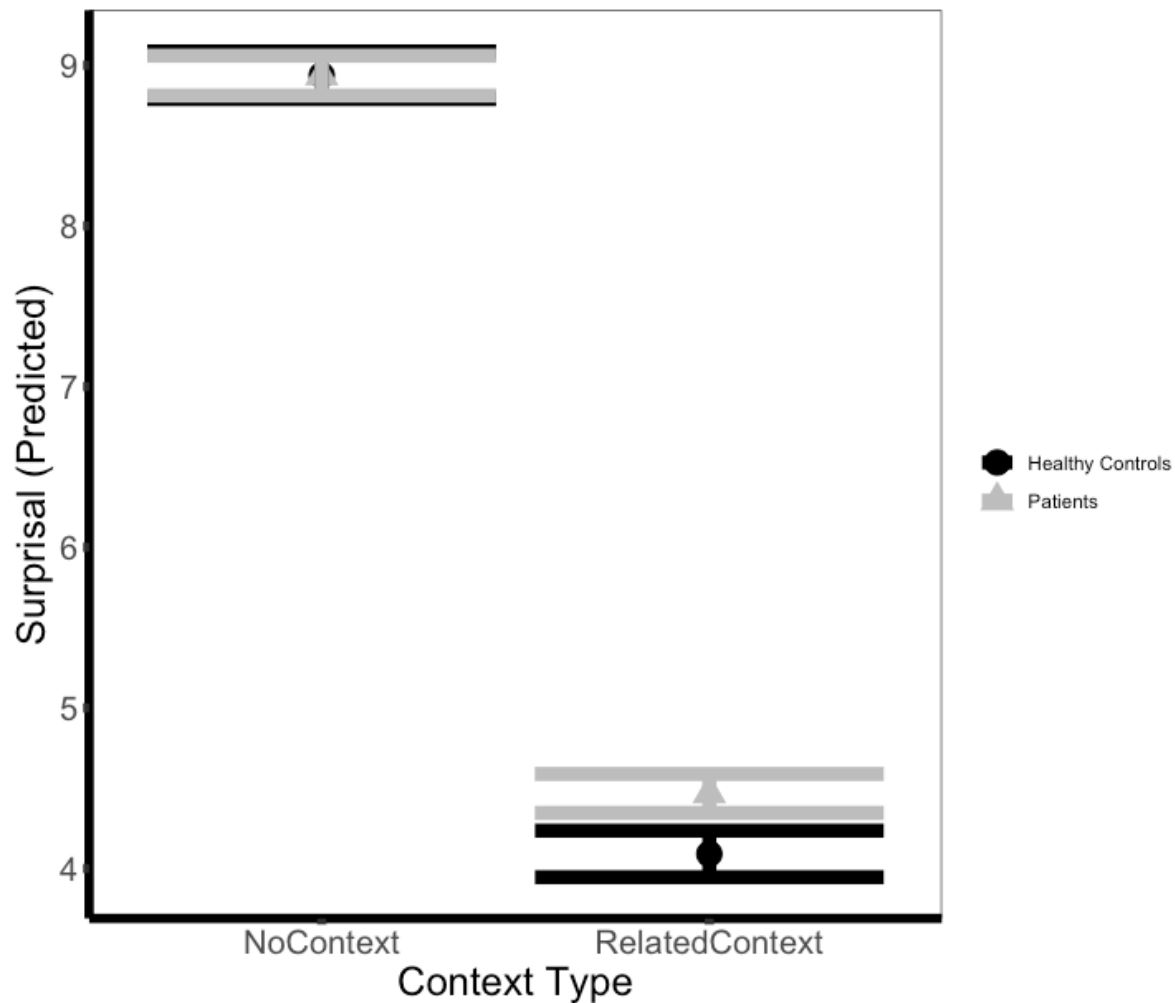


Figure 10. The effect of Group and Context Type on surprisal as predicted by the LMER. Error bars denote 95% CI.

Question 2: Can the overall difference in how patients and controls used context be explained by a general impairment in mobilizing semantic information appropriately?

While the previous analysis suggested that patients, overall, are not using contextual information to decrease surprisal in the same way as controls, it is unclear whether this difference is due to disruptions in the form of the speech output, or the content of the speech output. Disruptions in form describe differences in *how* ideas are transmitted (the structure of

speech, at various scales), while disruptions in content describe differences in *what* is being said. As such, disrupted content could stem from a complete inability to produce semantically cohesive speech (e.g. the “word salad” observed in some patients), or from atypical ideas which would be inherently less predictable to the model (e.g. a patient talking about a strange delusion or hallucination they’re having).

To probe whether the difference we observed in the previous analysis could be attributed to differences in form versus content, we computed word-by-word surprisal, using, as context, the actual words that the subject produced prior to the target word, but in randomized order. In this way, we perturbed the *structure* of their speech while leaving broader semantic field information (the content of their speech) intact. We then compared this new condition - SemanticContext - with the NoContext condition we probed previously. If our findings in the previous analysis stem from the content of patients’ speech, then we should expect to again see a significant interaction between Context Type (NoContext, SemanticContext) and Group, suggesting that patients overall fail to produce a semantic field that is cohesive and predictable, relative to healthy controls.

As in the previous analysis, there was no main effect of Group (see Table 10 for statistics). As expected, the main effect of Relatedness was significant and negative, such that in healthy controls, SemanticContext surprisal was less than NoContext surprisal. We found, however, that there was no evidence for a deficit in patients. That is, the interaction between Context Type and Group was non-significant and numerically small. Though we cannot say definitively, from this null effect, that our results are *not* driven by differences in the ability to generate predictable semantic content, we argue that this pattern of results suggests that disturbances in speech *structure* are more likely to be linked to atypically low lexical

predictability in patients. We return to this idea later on in analyses probing relationships with clinical ratings of formal thought disorder.

*Model Structure: Surprisal ~ Group*Context Type + SES* Context Type + Segment Length* Context Type + (1 + Context Type | Subject) + (0 + Context Type | Word)*

Effect of Group and Context Type (NoContext vs. SemanticContext) on Surprisal

	Estimate	SE	t	p	Sig.
Intercept	8.91	0.17	52.28	0.00	***
Group	0.00	0.10	0.05	0.96	
Context_Type	-1.11	0.19	-6.00	0.00	***
SES	0.03	0.05	0.60	0.55	
Segment Length	0.00	0.00	0.22	0.83	
Context Type*Group	-0.05	0.08	-0.64	0.52	
Context Type*SES_Z	0.03	0.04	0.68	0.50	
Context Type*Segment Length	0.00	0.00	1.04	0.30	

Table 10. Predictors of interest are shown in bold.

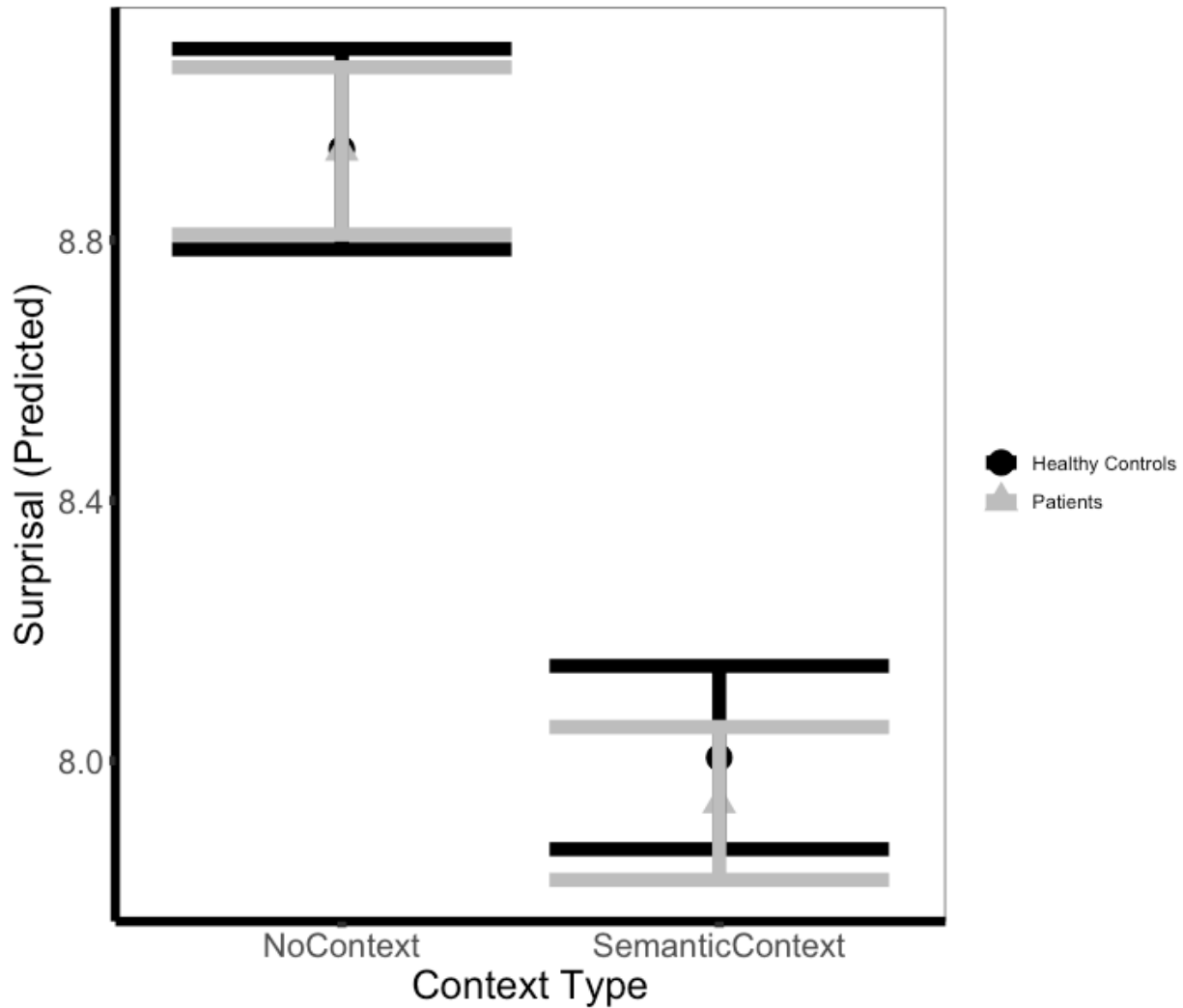


Figure 11. Effect of Group and Context Type on surprisal, as predicted by the LMER. Error bars denote 95% CI.

Question 3: Do patients exhibit a selective impairment in using global, versus local, context to produce upcoming words?

Having established that patients were engaging in task but failed to use context in the same way as healthy controls, we now turn to our main question of whether patients have a selective impairment in using global, versus local, context in order to produce speech.

As discussed in the introduction, lexical predictability is a holistic measure of how a word relates to its context. Surprisal, the negative log-transform of predictability, is another way

of quantifying this relationship; indeed, it decreases monotonically as predictability increases. However, surprisal is special in that it allows us to probe not only whether there is an *absolute* difference between patients and controls in how predictability changes as context builds up, but also whether the group difference in global predictability is *proportional* to the group difference in local predictability. That is, we expect differences in lexical predictability to increase as predictability itself increases, but a truly selective deficit in use of global context should manifest as a group difference in global predictability which is disproportionate to any group differences in local predictability - a prediction which is more easily tested in log, rather than linear, space (Gelman & Hill, 2006).

To test this prediction, we took each word in each sample of speech in this dataset and computed surprisal, for each word, with up to 50 different context lengths (i.e. with context truncated so that it included only the 1, 2, 3, ..., 50 words of context preceding the target word). These surprisal values served as the dependent measure in this analysis. Log Context Length (log-transformed to better fit assumptions of linearity), Group, and the interaction between the two were the predictors of interest. Based on previous work, we expected a main effect of Log context Length on surprisal, such that in healthy controls, surprisal would decrease as Context Length increased. However, more critically, we expected that surprisal would be overall greater (less negative) than in controls and that this difference between groups would increase with increasing Context Length (a positively-signed interaction between Group and Context Length). We note that, as discussed in the introduction, because Log Context Length varied *within* individual words, we here modeled by-word random effects (see Methods).

As expected, we saw a strong negative effect of Log Context Length on surprisal, such that, within controls, surprisal decreased as Log Context Length increased (see Table 11 for

statistics).⁶ There was also a main effect of Group, such that surprisal was greater in patients at the mean Log Context Length. Most importantly, however, there was a significant interaction between Group and Log Context Length, with a positive coefficient, such that surprisal decreased **less** in patients, relative to controls, given more and more context, consistent with our predictions.

*Model Structure: Surprisal ~ Group*Log Context Length + SES*Log Context Length + Segment Length*Log Context Length + (1 + Log Context Length | Subject) + (1 + Log Context Length | Word)*

Effect of Group and Log Context Length on Surprisal

	Estimate	SE	t	p	Sig.
Intercept	4.55	0.07	66.81	0.00	***
Log_Context_Length	-1.02	0.02	-60.47	0.00	***
Group	0.32	0.09	3.64	0.00	***
SES	0.07	0.04	1.81	0.07	.
Segment_Length	0.06	0.03	2.15	0.03	*
Log_Context_Length*Group	0.09	0.02	3.94	0.00	***
Log_Context_Length*SES	0.01	0.01	0.98	0.33	
Log_Context_Length*Segment_Length	0.06	0.01	7.61	0.00	***

Table 11. Predictors of interest are shown in bold.

⁶ Note that a 1.02 difference in surprisal is equivalent to approximately a 177% increase in predictability with each 1-unit increase in Log Context Length (e.g. from 1 to ~3 words, or from ~7 to ~20 words)

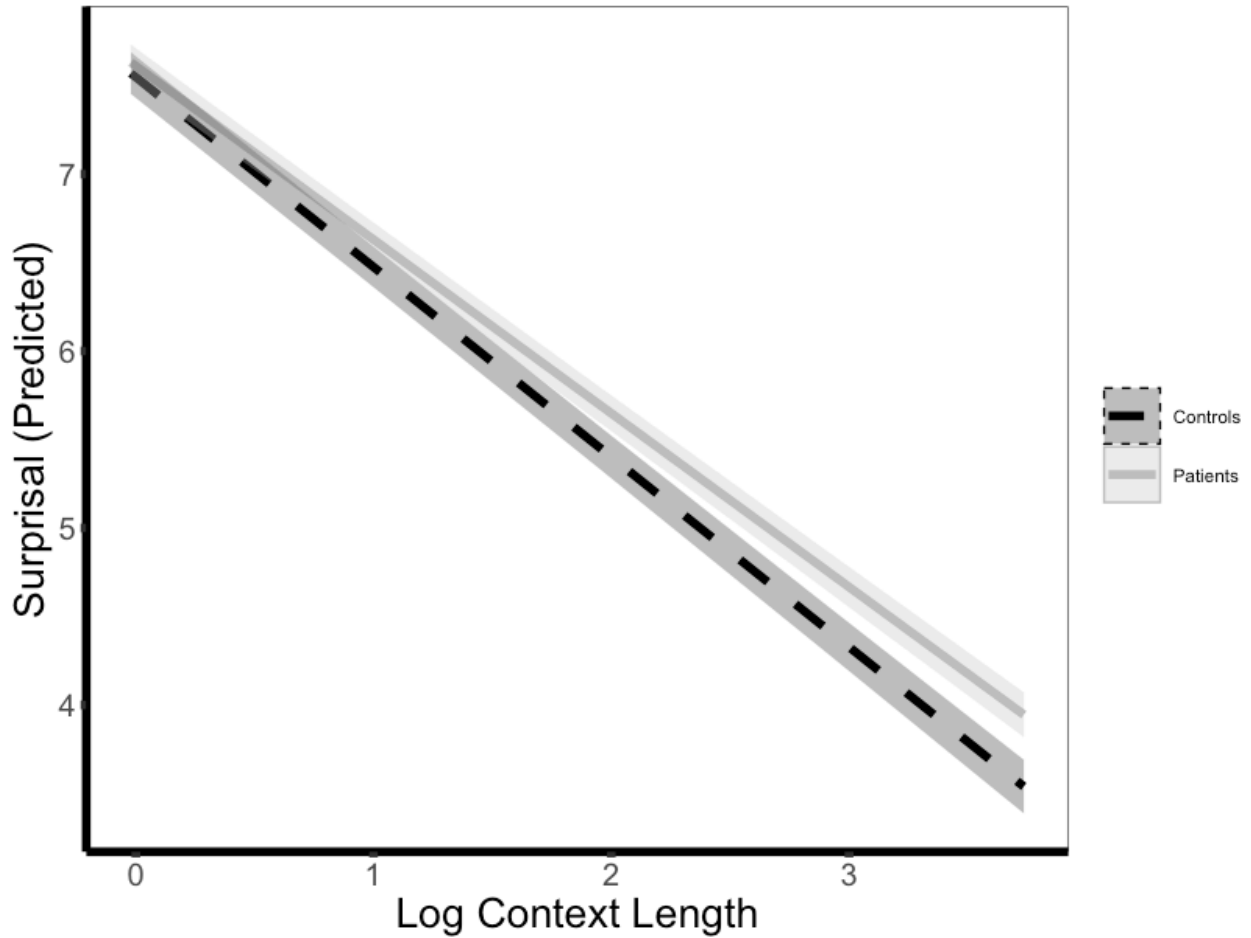


Figure 12. Effect of Group and Log Context Length on surprisal, as predicted by the LMER model. Error bars represent 95%

To confirm that the interaction we observed was due to a selective deficit in using global context, we followed up this interaction to determine whether it was driven by (a) decreased surprisal at more local windows in patients, (b) increased surprisal at more global windows in patients, or (c) both decreased local surprisal *and* increased global surprisal. To do this, we computed two summary created aggregate measures based on the extremes of our range of context lengths—*local surprisal* and *global surprisal*— by averaging, by word, the shortest 5 (1 to 5 words) context windows and the longest 5 (46 to 50) context windows, respectively. We tested for a main effect of Group on each of these measures on local and global surprisal

separately. The effect of Group on Global Surprisal was significant and positive, such that surprisal at these more global context windows was greater for patients than for controls (see Table 12 for statistics). In contrast, the effect of Group on local surprisal was non-significant (see Table 13 for statistics), such that there was no evidence for a difference in how patients versus controls were using local context. Together, these findings suggest that the interaction observed was primarily driven by differences in surprisal computed based on more global context lengths.

Model Structure: Global Surprisal ~ Group + SES + Segment Length + (1 | Subject); Local Surprisal ~ Group + SES + Segment Length + (1 | Subject)

Effect of Group on Global Surprisal

	Estimate	SE	t	p	Sig.
Intercept	3.88	0.07	51.83	0.00	***
Group	0.25	0.09	2.70	0.01	**
SES	0.10	0.05	2.10	0.04	*
Segment_Length	-0.05	0.06	-0.92	0.36	

Table 12. Predictors of interest are shown in bold.

Model Structure: Local Surprisal ~ Group + SES + Segment Length + (1 | Subject); Local Surprisal ~ Group + SES + Segment Length + (1 | Subject)

Effect of Group on Local Surprisal

	Estimate	SE	t	p	Sig.
Intercept	6.63	0.06	111.97	0.00	***
Group	0.12	0.08	1.56	0.12	
SES	0.04	0.04	1.17	0.25	
Segment_Length	-0.03	0.03	-0.94	0.35	

Table 13. Predictors of interest are shown in bold.

Are group differences in surprisal for local vs global context explained by impaired domain-general cognitive functions?

We next examined whether differences in use of local versus global context could be explained by domain-general cognitive functions as measured by the Digit-Symbol Substitution Test (DSST), the category fluency task, and the Trail-Making Test Part B (see Methods for details). We took the mean of these three scaled scores to form a composite measure, Overall Cognition. We then tested for an interaction between Overall Cognition and Log Context Length. We also, again, included an interaction between Group and Log Context Length in the model. If our results can be explained by differences in these cognitive functions, we should expect to see a significant interaction between Overall Cognition and Log Context Length over and above the interaction between Group and Log Context Length.

Again, we saw strong main effects of Log Context Length (see Table 14 for statistics) and Group, and an interaction between the two. However, both the main effect of Overall Cognition as well as the interaction between Overall Cognition and Log Context Length were non-significant, such that there was no evidence that differences in these scores could explain the group differences we observed.⁷

⁷ We note that when this analysis was repeated using the individual cognitive scores, the main effects and interactions remained non-significant, apart from a trending main effect of DSST score (Est. = -.08, SE = .05, $p = .09$).

*Model Structure: Surprisal ~ Group*Log Context Length +Overall Cognition*Log Context Length + SES*Log Context Length + Segment Length*Log Context Length + (1 + Log Context Length | Subject) + (1 + Log Context Length | Word)*

Effect of Overall Cognition, Group, and Log Context Length on Surprisal

	Estimate	SE	t	p	Sig.
Intercept	4.56	0.08	59.76	0.00	***
Overall_Cognition	-0.02	0.06	-0.34	0.74	
Log_Context_Length	-1.01	0.02	-55.16	0.00	***
Group	0.29	0.11	2.56	0.01	*
SES	0.09	0.05	1.80	0.08	.
Segment_Length	0.08	0.03	2.62	0.01	**
Log_Context_Length*Overall_Cognition	-0.01	0.01	-0.42	0.68	
Log_Context_Length*Group	0.08	0.03	2.89	0.01	**
Log_Context_Length*SES	0.01	0.01	1.16	0.25	
Log_Context_Length*Segment_Length	0.06	0.01	7.19	0.00	***

Table 14. Predictors of interest are shown in bold.

Question 4: Are differences in surprisal for local vs global context associated with positive thought disorder (within patients)?

Having established evidence for a selective deficit in using global context in patients, global-local, we then asked whether, within patients, this deficit could be specifically linked to positive thought disorder *within* patients. Thus, subsetting the data to only patients, we tested for an interaction between TLI Disorganization and Log Context Length (see Table 15 for statistics). Overall symptom severity as measured by the PANSS-8 total score, was included as an additional nuisance variable. There was, firstly, a significant main effect of TLI Disorganization, such that the difference in surprisal between groups was significant at the mean log context length. Indeed, we also found that the interaction between TLI Disorganization and Log Context Length interaction was significant and positive, such that the effect of Log Context Length on Surprisal was less negative in patients who were rated as having more severe speech disorganization.

We again followed up following up this interaction by looking at the effect of TLI Disorganization on our aggregate measures of Global and Local Predictability. In this analysis, we again showed a significant main effect of TLI Disorganization on Global Surprisal (See Table 16 for statistics) and a non-significant main effect of TLI Disorganization on Local Surprisal (see Table 17 for statistics).

*Model Structure: Surprisal ~ TLI Disorganization*Log Context Length + SES*Log Context Length + Segment Length*Log Context Length + PANSS-8*Log Context Length + (1 + Log Context Length | Subject) + (1 + Log Context Length | Word)*

Effect of TLI Disorganization and Log Context Length on Surprisal (Within Patients)

	Estimate	SE	t	p	Sig.
Intercept	4.91	0.08	62.80	0.00	***
Disorganization	0.07	0.04	1.67	0.10	
Log_Context_Length	-0.94	0.02	-47.77	0.00	***
Segment_Length	0.07	0.04	1.87	0.06	.
SES	0.10	0.05	1.91	0.06	.
PANSS-8_Total	-0.07	0.08	-0.91	0.37	
Log_Context_Length*Disorganization	0.03	0.01	2.63	0.01	*
Log_Context_Length*Segment_Length	0.06	0.01	6.42	0.00	***
Log_Context_Length*SES	0.00	0.01	-0.33	0.75	
Log_Context_Length*PANSS-8_Total	-0.01	0.02	-0.66	0.51	

Table 15. Predictors of interest are shown in bold.

Model Structure: Global Surprisal ~ Group + SES + Segment Length + (1 | Subject); Local Surprisal ~ Group + SES + Segment Length + (1 | Subject)

Effect of TLI Disorganization on Global Surprisal (Within Patients)

	Estimate	SE	t	p	Sig.
Intercept	4.06	0.05	74.20	0.00	***
Disorganization	0.09	0.04	2.62	0.01	*
Segment_Length	-0.04	0.07	-0.55	0.58	
SES	0.13	0.05	2.52	0.02	*

Table 16. Predictors of interest are shown in bold.

Model Structure: Local Surprisal ~ Group + SES + Segment Length + (1 | Subject); Local Surprisal ~ Group + SES + Segment Length + (1 | Subject)

Effect of TLI Disorganization on Local Surprisal (Within Patients)

	Estimate	SE	t	p	Sig.
Intercept	6.74	0.05	136.75	0.00	***
Disorganization	0.02	0.03	0.46	0.65	
Segment_Length	-0.03	0.03	-0.75	0.45	
SES	0.10	0.05	2.17	0.04	*

Table 17. Predictors of interest are shown in bold.

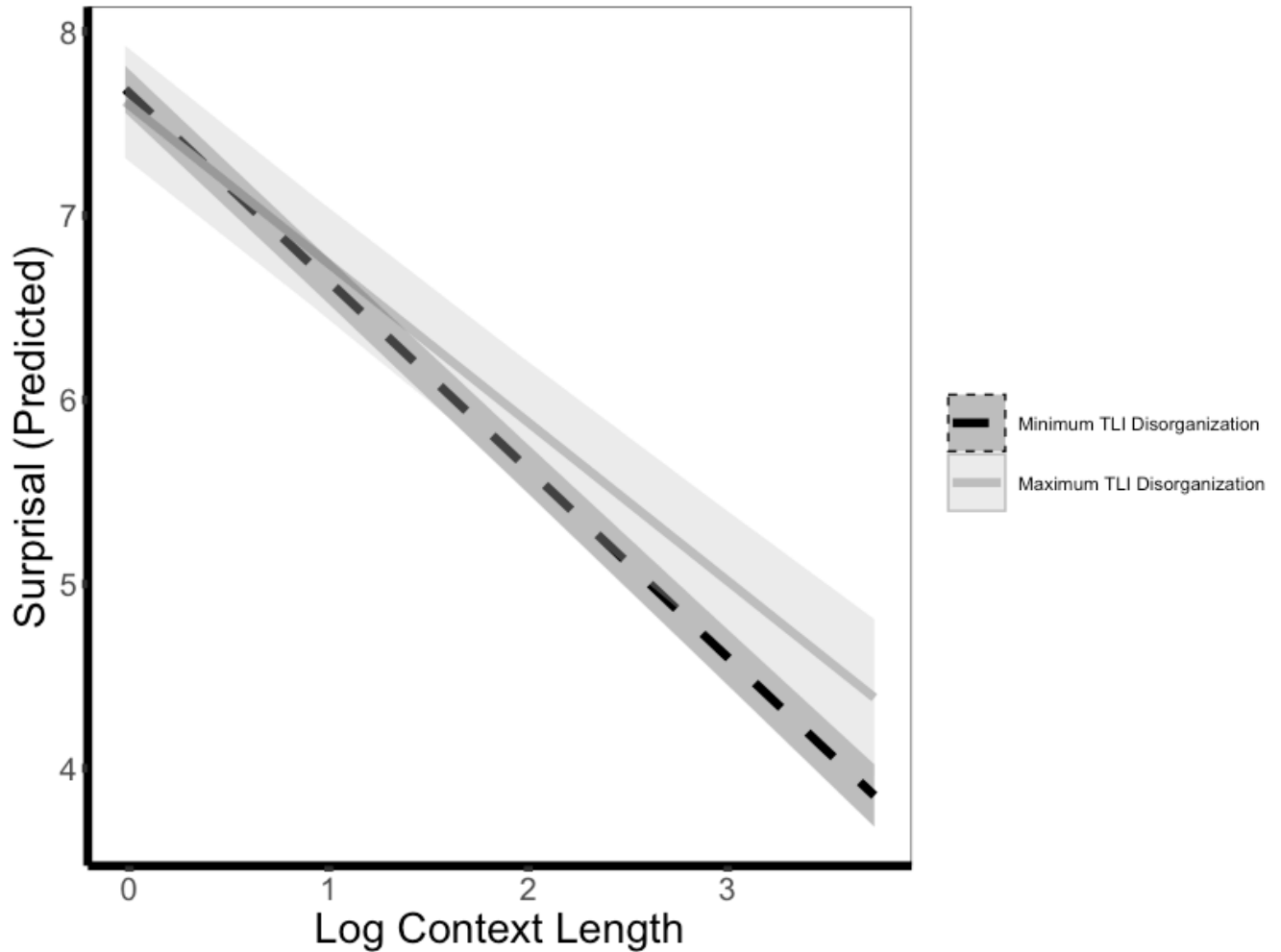


Figure 13. The effect of Log Context Length and TLI Disorganization on surprisal. Shaded areas represent 95% CI.

Finally, we tested whether differences in use of local versus global context could be linked to negative thought disorder. To do this, we looked for an interaction between TLI Impoverishment and Log Context Length, again accounting for Overall Symptom Severity. Here, we found that the main effect of TLI Impoverishment scores was non-significant (see Table 18 for statistics), and so was the interaction between Log Context Length and TLI Impoverishment scores. Note also that the coefficient and standard error for this interaction were numerically very small, relative to those observed in the previous analysis. Thus, there was no evidence for a relationship

between negative thought disorder and use of global versus local context.

*Model Structure: Surprisal ~ TLI Impoverishment*Log Context Length + SES*Log Context Length + Segment Length*Log Context Length + PANSS-8*Log Context Length + (1 + Log Context Length | Subject) + (1 + Log Context Length | Word)*

Effect of TLI Impoverishment and Log Context Length on Surprisal (Within Patients)

	Estimate	SE	t	p	Sig.
Intercept	4.93	0.08	64.23	0.00	***
Impoverishment	0.08	0.05	1.64	0.11	
Log_Context_Length	-0.93	0.02	-46.01	0.00	***
Segment_Length	0.08	0.04	2.25	0.02	*
SES	0.10	0.06	1.90	0.06	.
PANSS-8_Total	-0.06	0.08	-0.78	0.44	
Log_Context_Length* Impoverishment	0.00	0.01	0.31	0.76	
Log_Context_Length*Segment_Length	0.07	0.01	6.61	0.00	***
Log_Context_Length*SES	-0.01	0.01	-0.36	0.72	
Log_Context_Length*PANSS-8_Total	-0.01	0.02	-0.56	0.58	

Table 18. Predictors of interest are shown in bold.

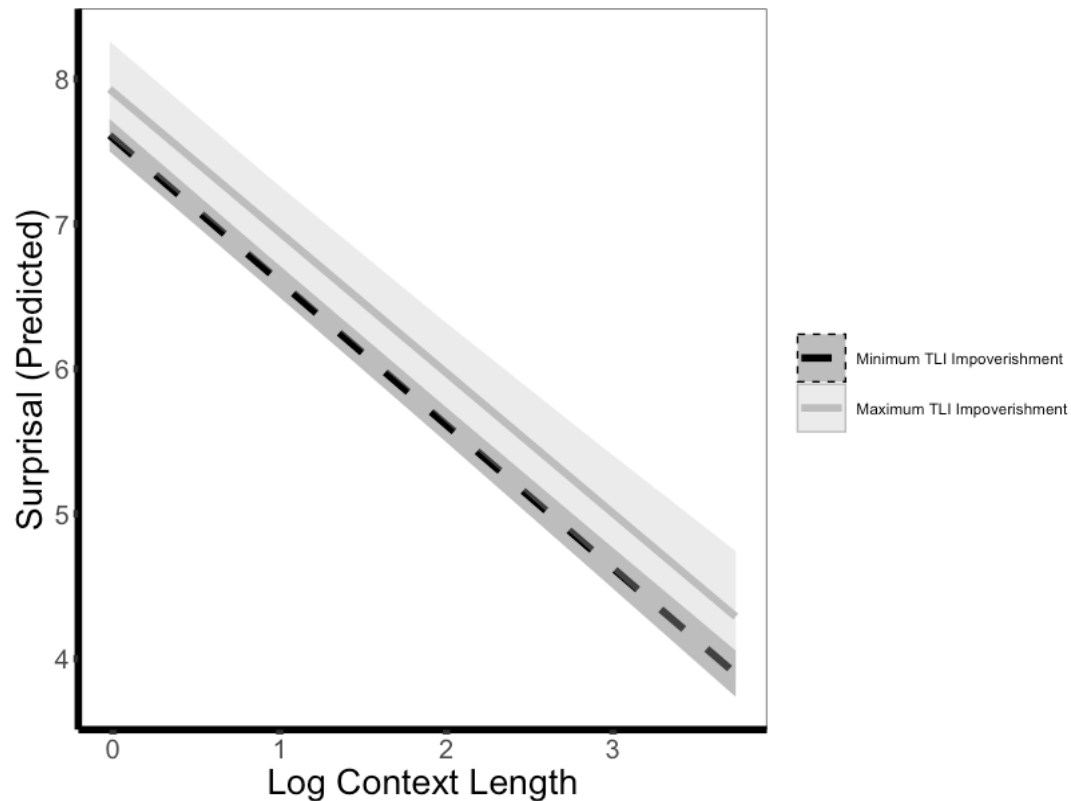


Figure 14. The effect of Log Context Length and TLI Impoverishment on surprisal. Shaded areas represent 95% CI.

Discussion

In the present study, we computed word-by-word lexical surprisal in samples of naturalistic speech produced by first-episode psychosis patients, as well as healthy controls. We found that there was an overall group difference in lexical surprisal, with surprisal being greater in patients than in healthy controls. More importantly, however, we found that this difference was driven by a selective deficit in incorporating context across longer temporal windows (i.e. global context), as evidenced by a larger difference in global surprisal across groups. In contrast, there was no evidence that use of more local context differed between groups. There was also no evidence that deficits in using global context could be explained by lack of attention to the task,

or by differences in overall cognitive functioning. Finally, we showed that impairments in using global context were linked specifically to speech disorganization, above and beyond overall symptom severity. Together, these findings bridge between clinical characterizations of thought disorder and a number of other literatures, including the psycholinguistic literature on language comprehension in schizophrenia and work across multiple domains evidencing selective deficits in processing more global information. The implications of our findings in light of these literatures are discussed below.

In general, effective communication is a complex task which requires taking into account many different constraints, including the mind of one's listener. Indeed, a large literature suggests that healthy adults regularly tailor their speech to be more understandable to the person they are speaking to (Clark & Murphy, 1982; Bell, 1984). In part, this means keeping track of what one has already said —i.e. the linguistic context— so as to present a message that is coherent from the listener's perspective. Because words that are more predictable are easier to process, it is to the speaker's benefit to make their speech more predictable to the listener (at least up to a point, as new information still needs to be transmitted), based on what they themselves have already said. Indeed, this is clearly evidenced by our data, in which surprisal decreases (i.e. predictability *increases*) given more and more context. This finding is in line the findings of Chapter 4, which show a strong negative effect of context length on surprisal and support the idea that healthy adults use context in relatively consistent, predictable ways which make comprehension easier on the listener. That is, instead of transmitting information more quickly by keeping surprisal high, healthy speakers appear to scaffold their upcoming speech by incorporating all context that is available to the listener. Indeed, the negative effect of Context Length in Surprisal in these data was extremely robust, and can even be clearly seen within

individual participants (see Figure A2, Appendix B).

In patients, Surprisal did still decrease given additional context, but not at the same rate as in controls. We saw that with longer windows of context, the difference in predictability between groups is disproportionately large, relative to the difference in predictability between groups given shorter windows of context, suggesting a specific, selective impairment in incorporating more global context to produce upcoming words. Indeed, there was no evidence that there was *any* difference in how patients and healthy controls made use of more local contexts. That is, patients appeared to be able to use context windows of up to at least 5 words as well as controls - enough to establish local dependencies (i.e. short phrases). Notably, these findings closely mirror the literature on context use during online language comprehension in schizophrenia, in which patients have been shown use local linguistic context (a clause or short sentence) to predict an upcoming target word just as well as controls, but fail to use context which is encoded at longer temporal scales (Ditman & Kuperberg, 2007; Swaab et al., 2013; Boudewyn et al., 2017). One question for future research will be whether, within individuals, impaired use of global context in comprehension necessarily translates to impaired use of global context in production, and vice versa.

We next asked whether this selective insensitivity to global context could be linked to speech disorganization. As discussed in the introduction, measures of global vs. local surprisal have face validity with Bleuler's conceptualization of "loosening of associations," which describes a tendency to produce locally associated words at the expense of global coherence (Bleuler, 1950). This phenomenon is now considered to be a key component of the disorganized speech that characterizes positive thought disorder (Andreasen & Grove, 1986). Indeed, we found that patients rated as having higher levels of speech disorganization, as measured by the

Disorganization subscore of the Thought and Language Index (Liddle et al., 2002), had greater deficits in using global context — even, importantly, when we accounted for overall symptom severity. In contrast, there was no evidence that negative thought disorder, as measured by the TLI Impoverishment subscore, was associated with either surprisal more generally, or a difference in global vs. local surprisal. Together, these findings suggest that the face validity of global/local surprisal for measuring positive thought disorder translates into an objective and easily measurable index of disorganized speech. Moreover, there was some evidence, though weak, that even patients with low disorganization ratings may differ from healthy controls in their use of global context. More research is needed to fully understand this relationship, but this may be a hint that global vs. local surprisal is a quite sensitive measure, picking up on speech impairments that are not detectable in clinical ratings. Such a measure could be developed into a powerful biomarker for psychosis, as discussed in Corcoran et al. 2020.

Thus, we identified in first-episode psychosis patients a selective deficit in using global context which is clearly linked to speech disorganization and which is not explained by inattention to task or linked to domain-general cognitive functioning. An open question, then, is what *can* explain such an impairment. One possibility, which has received much attention recently, is the presence of disruptions in the hierarchical generative circuits that govern language processing (Brown & Kuperberg, 2015; Palaniyappan & Venkatasubramanian, 2022). Such a framework, which is based on principles of Bayesian updating, posits that, in healthy adults, language production (and other cognitive domains, see, e.g. Adams et al., 2013, for discussion) is supported by information flowing up and down a hierarchy of representations, incrementally updating the speaker’s “mental model” of their message and the consequences of that message. Within this hierarchy, local linguistic context would be presumed to primarily

contain low-level information—that which spans shorter temporal scales like local syntactic dependencies—whereas more global context would be presumed to primarily contain “high-level” information—that which spans longer temporal scales, like discourse structure. It has been suggested that a failure to appropriately propagate prediction error in a top-down fashion (perhaps as a result of highly imprecise priors, see Adams et al., 2013, for discussion) would result in speech that is overly determined by local information, at the expense of global coherence (Brown & Kuperberg, 2015) — in other words, speech that looks very much like what patients produced in this study.

Of course, one critical limitation of *this* study is that it cannot tell us very much about the specific neurocomputational underpinnings of this failure to use global context. This is because large language models like GPT are not generally designed with biological plausibility in mind and as such, lack the robust top-down feedback connections that are, by many researchers, presumed to be an integral part of how the brain does language (Farmer, Brown, & Tanenhaus, 2013; Kuperberg & Jaeger, 2016; Kuperberg, 2021). Based on GPT-3’s performance on language tasks (Brown et al., 2020), and our work in Chapter 4, it is clear that for the most part, GPT can approximate human intuitions about language very well, making it suitable for descriptive characterizations of language production, which can in turn help us begin to understand, broadly, the types of mechanisms underlying language impairments. However, it is likely that the model acquires and uses these intuitions in very different ways than humans do; for this reason, it is impossible for these data to make specific claims about how such mechanisms manifest in the brain.

To do this, additional tools will be needed, which may include computational models which are more biologically plausible. Specifically, we suggest that modeling disordered

language within a predictive coding architecture could be a natural next step. Predictive coding refers to a particular computational algorithm and architecture which describes how information is transmitted amongst cortical areas, allowing the brain to perform probabilistic inference (Friston, 2005; Clark, 2013; Spratling 2016). This transmission of information relies on propagation of predictions and prediction error, and resultant updating of internal representations; as such, the predictive coding account of language is one implementation of a hierarchical generative framework, which we discuss above. However, predictive coding in particular has received recent attention for its ability to simulate various perceptual and cognitive phenomena (e.g., Spratling, 2016), including effects of context on language processing (Nour-Eddine et al., under review). Importantly, atypicalities in predictive coding have also been proposed to provide a holistic, biologically plausible explanation for many of the symptoms of psychosis (Adams et al, 2013; Sterzer et al., 2018). Thus, an important goal of future studies should be to understand whether, and how, selective deficits in using global context during language production can be modeled in terms of aberrant predictive coding.

Neurally, our findings are also consistent with work suggesting that neurons encode shorter temporal receptive windows (also known as ‘intrinsic neural timescales’) in schizophrenia (Uscătescu et al., 2021; Uscătescu et al., 2023). In healthy adults, it has been shown that different brain regions track and integrate context over different timescales (Hasson et al., 2008; Lerner, Honey, Silbert & Hasson, 2011). Studies in schizophrenia have shown clear atypicalities in the temporal receptive window, though the directionality of these atypicalities is somewhat inconsistent, with some finding longer temporal windows (Northoff et al., 2021; Lechner & Northoff, 2023) and some finding shorter temporal windows (Uscătescu et al., 2021; Uscătescu et al., 2023) in patients. To our knowledge, no studies in schizophrenia have yet

investigated the length of the temporal window in language specifically; however, given our findings, we would predict reduced timescales relative to healthy controls. These differences may be even greater for patients who present with greater speech disorganization.

In summary, the present study provides evidence that patients in the acute phase of schizophrenia exhibit a selective deficit in using global context which indexes clinical, and potentially sub-clinical, speech disorganization. Thus, the measures computed here have important clinical relevance in addition to touching on more theoretical questions. Indeed, these findings hint at the type of cognitive mechanisms that may underlie disordered language production in schizophrenia, and raise questions regarding the specific neurocomputational correlates of insensitivity to global context. We suggest these findings scaffold multiple lines of future research, ranging from the development of highly sensitive language-based biomarkers to the simulation of disordered speech through powerful, biologically plausible computational architectures like predictive coding.

Chapter 6. General Discussion

In this thesis, we report findings from three studies examining language production, under naturalistic and semi-naturalistic conditions, in healthy controls and people with schizophrenia. Together, these studies paint a picture of selectively impaired language production in schizophrenia. In this final chapter, we will synthesize the conclusions that can be drawn from these findings and discuss their broader implications within the wide-ranging and varied field of schizophrenia research.

Summary of Findings

In Chapter 3, we used a collaborative picture-naming paradigm to probe a phenomenon in language production known as lexical alignment. As we discuss in this chapter, lexical alignment, in healthy adults, is thought to be driven by both bottom-up priming mechanisms (Pickering & Garrod, 2004) and top-down mentalizing (Branigan et al., 2011). Though previous literature has shown substantial evidence for mentalizing deficits in schizophrenia (for review, see Brune, 2005), there has been very little work on the degree to which those deficits affect language production, particularly in communicative contexts. Indeed, the study we report in Chapter 3 was the first study, to our knowledge, to examine lexical alignment in people with schizophrenia. We showed that, while patients lexically aligned with their conversation partners just as well as healthy controls, alignment in patients was **more** associated with lexical production priming and **less** associated with mentalizing than in healthy adults. In other words, in patients, top-down mentalizing failed to drive lexical alignment, but these deficits were compensated for by increased use of bottom-up mechanisms. These findings are in line with previous work in comprehension showing that patients fail to use broad goal-relevant context in

a top-down fashion to predict upcoming words (Barch et al., 1996; Barch & Ceaser, 2012; Sharpe, Weber, & Kuperberg, 2020), whereas use of bottom-up mechanisms is largely intact (Barch et al., 1996; Blum & Freides, 1995; Chapin et al., 1992; Ober et al., 1995; Vinogradov et al., 1992).

In Chapters 4 and 5, we took a different approach, shifting from a tightly controlled paradigm focused on collaborative aspects of communication to using computational measures to explore language output under more naturalistic conditions. In both chapters, we show that, for healthy controls, context length (the number of words of linguistic context preceding a “target” word) is strongly, negatively associated with lexical surprisal (a measure of how *unpredictable* a word is based on its context). Together, these chapters provide evidence that healthy adults make use of context in predictable, normative ways. Chapter 5 builds on these findings by asking how these normative patterns of using context may be disrupted in schizophrenia. We show, using a novel, theory-driven approach, that surprisal also decreases, given more and more linguistic context, in patients, but it is to a lesser extent than in healthy controls. This finding suggests a selective deficit in the use of global linguistic context for people with schizophrenia. Though this is a new and exciting finding in language production, it is in keeping with prior work in language *comprehension* (Ditman & Kuperberg, 2007; Swaab et al., 2013; Boudewyn et al., 2017) and also work in other perceptual and cognitive domains, e.g. vision (Johnson, Lowery, Kohler, & Turetsky, 2005). Thus, our findings lay important groundwork for explicitly linking language production with the broader symptomatology of schizophrenia. Notably, we also found that the severity of the selective deficit in using global context was linked to subjective ratings of positive thought disorder, supporting the face validity of this construct with Bleuler’s conceptualization of “loose associations” (Bleuler, 1950).

Broader Implications

How are we to synthesize the findings of these studies? One way to situate these findings within a common framework is to consider the role of top-down processes in language production. In Chapter 3, of course, we explicitly contrast top-down mechanisms (mentalizing, or theory of mind) with bottom-up mechanisms (priming) by encouraging lexical alignment, which is thought to be dependent on a combination of these two mechanisms. As discussed above, we saw that, as expected, healthy controls appeared to use a combination of low-level (priming) and high-level (mentalizing) representations to align their lexical choices to their partner's speech. In contrast, within the schizophrenia group, bottom-up priming mechanisms were increased relative to healthy controls, whereas top-down mentalizing was impaired.

In Chapter 5, however, we probed the contrast between top-down and bottom-up mechanisms more implicitly, as a function of the size of the temporal window represented by context length. That is, longer temporal windows are more likely to capture higher-level representations (e.g. message-level representations) which are maintained over longer timescales whereas shorter temporal windows are more likely to capture low-level representations (e.g. local syntactic dependencies) which are maintained over shorter timescales. Language, both hierarchical and highly interactive, is thought to critically depend on a proper balance between the information flowing “upward” from low-level representations and the information flowing “downward” from higher-level representations. Thus, while these types of mechanisms cannot be directly distinguished by our experimental design, the fact that use of global context appears to be specifically impaired hints at a failure of top-down processing.

In this way, the findings reported in this thesis, together, begin to link language

production with a larger literature on the neurocognitive underpinnings of schizophrenia, in which some have posited impaired top-down processing as a fundamental deficit driving many aspects of the disorder (see Brown & Kuperberg, 2015, for review). One major difference between these studies, however, is that in Chapter 3, patients were able to compensate for their lack of top-down processing with increased bottom-up mechanisms, resulting in performance that was the same as controls'. In Chapter 5, however, there was no evidence for increased use of low-level information: there was a clear group difference in overall surprisal, and local surprisal was not greater in patients than in controls. We believe this may be linked to the inherently greater complexity of the task posed in Chapter 5: patients were asked to describe pictures of scenes, with multiple people and interpretations, rather than to simply name pictures of single objects. This additional complexity likely translates to higher task demands. That is, during more fluent, naturalistic speech, even when only accounting for a single word of context, multiple constraints must be taken into account which are not at play in a simple naming task (e.g. syntactic constraints). As such, relying on very simple bottom-up mechanisms like lexical priming may not be enough to reduce surprisal to the same extent as in controls. Indeed, though Chapter 3 brings some elements of naturalness to the experimental design, the communicative task still does not approach the level of complexity involved in natural conversation. Thus, failures of top-down processing may cause much more noticeable deficits in the context of a real-world communication.

Future Directions

As we have discussed, this thesis advances the field's understanding of language production in schizophrenia by applying novel psycholinguistic and computational approaches

whilst incorporating elements of naturalistic communication, thus beginning to fill a number of gaps in the literature, which we discuss in detail in Chapter 2. However, there is still much to be done toward gaining a more comprehensive picture of how and why language production and comprehension are affected in schizophrenia. Indeed, our research highlights at least several important open questions that remain.

As discussed above, though we show that at least under some conditions patients' tendency to lexically align with a conversation partner is the same in patients as in healthy controls; however, it remains unknown whether lexical alignment (as well as other types of linguistic alignment, or other manifestations of audience design) is impaired in patients under more naturalistic conditions. Indeed, scaling up the complexity of the sort of task we have created may provide a fruitful avenue for future research; we would predict that as complexity increases, lexical alignment may begin to decrease in patients relative to healthy controls.

Supporting this hypothesis is some preliminary data from a study by Achim and colleagues (2023), which used a similar but somewhat more complex cooperative “game”, in which subjects were asked to describe a layout of pictures so that their partner could replicate it. In this task, patients were less likely than healthy controls to adapt their specificity of how they described the pictures in line with their partner's expected knowledge. Thus, it is plausible that the severity of deficits in lexical alignment and other types of audience design may be much greater in the context of real-world communication. In contrast to our study, however, it was not possible here for patients to use lexical production priming to compensate for impaired top-down mechanisms; thus it remains unclear whether patients' atypicalities in using top-down mentalizing mechanisms are *always* detrimental under more complex, naturalistic conditions. One can easily imagine, however, leveraging this type of task to look at lexical alignment as

well, using more back-and-forth between partners, in order to get a better sense of how top-down and bottom-up mechanisms interact in schizophrenia under greater task demands.

It will similarly be important to expand our understanding of how computational measures like lexical surprisal behave under more naturalistic, dialogic conditions. Though it is likely that GPT-2 performs rather poorly on dialogues versus on monologues, recent advances in the GPT series (with GPT-3 now available to the public via API and GPT-4 implemented via ChatGPT) may facilitate better estimates of dialogic predictability, particularly with additional fine-tuning of the model. These estimates would allow for more nuanced descriptions of how patients' speech relates to that of their partner, perhaps capturing aspects of positive thought disorder which are not detectable in monologues (e.g. the tendency to respond tangentially to questions; Andreasen, 1986). On the cognitive side, looking at predictability in dialogue may provide one way of linking production and comprehension deficits in conversation. For example, there is ample evidence for aberrant conversational turn-taking behavior in people with schizophrenia (Mueser, Bellack, Douglas, & Morrison, 1991; Colle et al., 2013) and in those at clinical high risk for psychosis (Sichlinger, Cibelli, Goldrick, & Mittal, 2019). Though in healthy adults, prediction of the end of a partner's turn appears to rely on lexicosyntactic cues (de Ruiter, Mitterer, & Enfield, 2006), it may be that this type of predictive processing is impaired in schizophrenia. Moreover, the degree to which patients are able to incorporate such cues into their *own* speech is unknown. If patients exhibit impairments in producing conversation turns which end predictably, and if such impairments are associated, on an individual level, with reduced ability to predict the end of their partner's turn, this could be interpreted as evidence for a more general deficit in generating predictions about conversational structure based on linguistic context. Similar investigations manipulating context length, like that which we performed in

Chapter 5, may additionally shed light on how the temporal window across which such cues unfold can affect patients' sensitivity to them.

Just as it is important to consider linguistic and conversational context, as well as factors like mentalizing that may drive audience design, it will be equally important to understand how other types of context affect how patients produce speech. For example, in Chapter 5, there was at least one very important piece of context - the visual context of the picture the subject was describing - that the model was unable to take into account. Therefore, while speech might be quite predictable based only on the words produced previously, the same words may be less predictable to someone who is looking at the picture they are describing (e.g. if the patient describes a character in an atypical way, or goes off on a tangent that is relevant to the topic of the picture but is not *describing* the picture, etc.). Thus, we may in fact be *underestimating* the extent of patients' deficits in using high-level context to produce upcoming words. Indeed, the question of how patients use visual information is particularly interesting given a large literature on selective deficits in schizophrenia for using more global aspects of visual context (see Butler, Silverstein, & Dakin, 2018, for review), as well as some evidence that high-level visual cues are not taken into account during online language comprehension (Rabagliati et al., 2018). There are several ways in which one could begin to understand how these types of information, which may not be encoded in linguistic context, are borne out (or not) in patients' speech. Topic modeling, for example, if validated on a separate set of controls' descriptions, could give a sense of how specific higher-level topics are distributed in the language output produced by patients versus by healthy controls. Similarly, a large language model like GPT-3 might be trained, or fine-tuned, to adjust its expectations about what the language output for this specific task should look like, based on a validation sample of healthy controls' speech. One might expect the contrast between

global and local surprisal to be even more drastic given this additional context.

In Conclusion

As discussed in the section above, this thesis lays groundwork for multiple avenues of future research, all of which can help us gain a clearer picture of the specific atypicalities that drive impaired language production in schizophrenia. Filling in this picture is important for a number of reasons. First, studying impaired language production in patients has implications for how we understand language more generally. When and why is prediction strictly *necessary* for effective and efficient communication, and when is it possible to compensate using bottom-up mechanisms? To what extent do comprehension and production draw on similar mechanisms and representations? How does a *balance* between top-down and bottom-up activity scaffold acquisition and maintenance of lexicosemantic representations? To what extent must different aspects of language processing draw on domain-general cognitive functions? Developing our understanding of how sensitivity to different types of linguistic- and non-linguistic context is borne out in schizophrenia speaks to these questions as well as many other fundamental questions about how human brains support language comprehension and production. More importantly, however, developing a better understanding of language impairment in schizophrenia the potential to fundamentally improve the lives of people with this disorder. More sensitive, comprehensive descriptive measures may aid in improved prediction of conversion to psychosis, allowing for early intervention; in helping clinicians track symptoms and progress; and in predicting patient outcomes. They may also provide meaningful targets for the assessment new drugs and treatment protocols. Moreover, understanding how and why insensitivity to different types of context impairs language production in real-world settings holds great promise

for developing targeted psychosocial rehabilitation programs which may alleviate impairments in social and occupational functioning. In turn, this may result in overall improved quality of life for patients. Finally, piecing apart the specific neurocognitive mechanisms driving impairments in using global context during language production, and how these impairments relate to atypicalities across other domains, may provide a basis for development of powerful cognitive remediation therapies, with the potential to improve more treatment-resistant symptoms and allow people with schizophrenia to live their lives more independently.

Appendix A

We both say tomato: Intact lexical alignment in schizophrenia and bipolar disorder

1. Recruitment and Characterization of Participants

Thirty-two outpatients with a DSM-IV diagnosis of a schizophrenia/schizoaffective disorder, and 32 outpatients with a DSM-IV diagnosis of bipolar disorder, were recruited from McLean Hospital. Of the patients in the schizophrenia group, 20 were diagnosed with schizophrenia and 12 were diagnosed with schizoaffective disorder. Of the patients in the bipolar group, 29 were diagnosed with bipolar I (15 with psychosis) and three were diagnosed with bipolar II. Diagnoses were confirmed using the Structured Clinical Interview for DSM-IV (First et al., 2002).

Patients' symptoms were assessed using the Positive and Negative Symptom Scales (PANSS; Kay et al., 1987), the Multnomah Community Ability Scale (MCAS; Barker et al., 1994), and the Young Mania Rating Scale (YMRS; Young et al., 1978). Assessments occurred within eight weeks of the experiment session. Assessments were completed by a single researcher who underwent extensive training in the administration of these scales and established high inter-rater agreement with scorers in other studies.

All participants were native speakers of American English (learned English before age 5, while growing up in the United States), who considered English to be their primary language.

2. Details of Cognitive Assessments

Verbal working memory, digital working memory, and executive function were measured following the procedures delineated in the Brief Assessment of Cognition in Schizophrenia

(BACS; Keefe et al., 2004, 2008).

Specifically, to assess verbal working memory, the experimenter read a list of fifteen words after which the participant was asked to recall all the words they remembered. The score was calculated as the number of correctly recalled words over the course of five trials. To assess digit span, the experimenter read a list of numbers and the participant then repeated them back in order from lowest to highest. The size of the list of numbers increased from two items to eight items, and the score was calculated as the number of correctly recalled items overall. Executive function was measured using the “Tower of London” task; in each trial, participants were shown two pictures of colored balls on a set of pegs: a “before” and an “after” picture. The participant’s task was to make determine the *fewest* number of moves it would take to make the arrangement in the before picture look like the arrangement in the after picture. The score was calculated as the number of correct responses over 21 trials.

3. Picture Naming Game: Development of Stimuli

All pictures were color photographs of objects, animals or people, which were stripped of their original surroundings and presented on a white square superimposed on a black background. We developed three sets of picture stimuli:

3.1. Dual-name pictures used to examine lexical alignment

These constituted 20 pictures that could be named with either a “preferred” or a “dispreferred” name (e.g. “bunny”/”rabbit”). To develop these stimuli, we began with a set of 45 pictures of objects that could be named with more than one name, selected from two previous studies (Branigan et al., 2016; Kuperberg et al., 2018), and carried out two online (Amazon

Mechanical Turk) norming studies, using participants between 30 and 60 years of age, all of whom reported American English as their first language, to narrow these down.

In the first norming study, 50 participants were presented with 60 photographs: the 45 items that were thought to elicit multiple names, and 15 filler items that were thought to only elicit a single name (e.g. “apple”). The order in which the pictures were presented was randomized between participants. For each picture, participants were asked to provide the first name that came to mind, and up to three alternative names. Eight participants were excluded based on their age (< 30). Data from 42 eligible participants (mean age 38.64 years old (SD: 7.98, range 30-57); 19 males, 21 female, 2 unknown) were included in analyses. Experimental items were selected for further pretesting if the most frequently produced name (“preferred name”) was provided by at least 60% of the participants. Thirty-two items matched these criteria (mean: 87.58%, SD 10.31%). Alternative names (“dispreferred names”) for these pictures were selected in two ways: 1) If the picture was described with the preferred name by 100% of the participants, we selected an alternative name that we thought was a good description of the picture (e.g. “trophy” was used by 100% of the participants, we selected “award” as alternative); 2) If participants had provided different names for the picture as the first name that came to mind, we picked the second most frequent name to be the alternative name (e.g. “rabbit” (86%), “bunny” (11.6%)).

In the second norming study, we checked the acceptability of the alternative/dispreferred names of these 32 selected pictures in an additional 30 participants (mean age: 39.6, SD: 7.5, range 31-59; 14 female) who hadn’t participated in the first norming study. These participants were presented with picture-name pairs and asked how acceptable they thought the names were for the pictures. They indicated their opinion on a scale from 1 to 7, where 1 was completely

unacceptable and 7 was completely acceptable. For each of the 32 experimental items, half of the participants saw the picture with the preferred name and half of the participants saw the picture with the dispreferred name. No participant saw the same item twice and conditions were counterbalanced between lists. To enhance variation and to make sure that participants were paying attention, we included 10 filler items that were presented with a clearly unrelated name (e.g. a picture of a sneaker with the name “mouse”) and 10 filler items that were presented with the name of the category to which the item belonged (e.g., a picture of a trumpet with the word ‘instrument’). The order in which the items were presented was randomized for each participant.

For all experimental items, preferred names had an acceptability rating of 6.5 or higher (mean 6.92, SD 0.09). Items were included if the dispreferred name had a rating of 5 or higher (out of 7). There were 27 items for which the dispreferred name received a 5 or higher (mean for included items: 6.31, SD = 0.52). From this set of 27 items, items were excluded that had a dispreferred name which was a compound word (e.g. “merry-go-round”) and/or for which the dispreferred name did not appear in the SUBTLEX database (via English Lexicon project: <http://lexicon.wustl.edu/>). We then picked the 20 items that had the highest acceptability scores for their dispreferred names. The preferred and dispreferred names for the final set of 20 dual-name pictures are shown in Table A1.

3.2. Single-name low-frequency pictures used to examine lexical priming

These constituted 40 pictures that had only one name, which was matched in frequency with the dispreferred name of the multiple-name items described above. To develop these 40 stimuli, we first looked up the SUBTLEX frequency (Brysbaert & New, 2009) of every dispreferred name, and used the same database to select at least two possible control items that

were matched in frequency. We selected words that were easily depicted with a simple photograph. Google was used to find freely usable pictures that depicted the objects described by the words. We started with an initial set of 66 pictures, and we carried out norming, using the same procedure as the first norming study described above, to ensure that that objects on these pictures were recognizable and elicited the names expected (but no other names). In total, we tested 66 possible control pictures (and 30 filler items), divided across three experimental lists. We recruited 30 participants for each list. For list 1, data from 27 participants (mean age 41.4 years (SD: 9.75, range: 30-60) 10 male, 17 female) were included in analyses. For list 2, data from 29 participants (mean age 40.24 years (SD: 8.61, range 30-59) 20 male, 9 female) were included in analyses. For list 3, data from 28 participants (mean age 37.93 years [SD: 8.78, range 30-59] 13 male, 15 female) were included in analyses. We excluded items which were named with the same name by less than 60% of the participants. Of the remaining low frequency items, for each dispreferred name we selected the two matching control items for which most participants provided the target name. On average, 92.22% (SD: 8.04) of all participants used the target name for the control items.

Dual-Name Pictures	Single-Name Low Frequency Pictures		
Preferred Name	Dispreferred Name	Control Item A	Control Item B
Axe	Hatchet	Octopus	Cherries
Boat	Yacht	Candle	Ankle
Bucket	Pail	Ostrich	Caterpillar
Cat	Kitty	Cigarette	Belt
Couch	Sofa	Tomato	Pigeon
Curtains	Drapes	Rake	Zipper
Sword	Katana	Xylophone	Teabag
Footprint	Footstep	Pomegranate	Pinecone
Frying pan	Skillet	Eyelashes	Calculator
Glasses	Spectacles	Mammoth	Milkshake
Helicopter	Chopper	Ham	Chef
Motorcycle	Motorbike	Kiwi	Tongs
Pillow	Cushion	Mushroom	Pineapple
Rabbit	Bunny	Tiger	Socks
Rock	Stone	Bell	Soldier
Rug	Carpet	Eagle	Fingerprints
Sled	Toboggan	Toucan	Tweezer
Suitcase	Luggage	Nails	Pumpkin
Sweater	Turtleneck	Treadmill	Clipboard
Trophy	Award	Cigar	Beard

Table A1. Preferred and dispreferred names of dual-name pictures, used to examine lexical alignment, and low-frequency names of single-name pictures used to examine lexical priming (matched for frequency with dispreferred names).

3.3 High frequency single-name pictures: Fillers

We also selected a set of 120 high frequency single-name items: pictures for which there was only one acceptable name. Frequency was checked using the SUBTLEX corpus. Because these trials were of no experimental interest, we did not conduct a norming study to confirm the acceptability of the names.

4. The Collaborative Picture Naming Game: Full Design and Counterbalancing

As outlined in the main manuscript, participants played a picture naming game with the experimenter in which the participant and the experimenter alternated naming pictures on a screen. *Production* trials, in which the participant named a picture, are described in the main text. Below, we outline the distribution of *comprehension* trials, in which the experimenter named the picture and the participant listened. In the case of the dual-name pictures and the single-name low frequency pictures, the *n*th comprehension trial served to “prime” participants’ *n*+3th production trials.

Dual-name pictures (20 trials). The experimenter named all 20 dual-name pictures prior to the participant in comprehension trials, with two intervening trials between the prime (comprehension trial) and the participant’s target (production trial). For half of these trials, the experimenter used the preferred name; for the other half, they used the dispreferred name.

Single-name low frequency pictures (20 trials; 2 conditions: repeated, unrepeated). The experimenter named 10 pictures with low frequency names. These were the primes for the

repeated low frequency single-name pictures. There were two intervening trials between these comprehension trials and the corresponding repeated production trials.

Unrepeated high frequency fillers (50 trials). The experimenter named 50 pictures with high frequency names, none of which had been shown before, and none of which were later repeated.

Repeated high frequency fillers (80 trials). The experimenter named 30 pictures with high frequency names. All of these pictures had been previously named by the participant, in production trials.

Counterbalancing and creation of lists

The two conditions of the dual-name picture production trials (*preceded by dispreferred name, preceded by preferred name*), and the two conditions of the single-name low frequency production trials (*repeated, unrepeated*) were counterbalanced across lists. This ensured that, in the dual-name production trials, while no individual participant saw the same dual-name picture in more than one condition, across all participants, a given picture was preceded by the dispreferred name about half the time and the preferred name about half the time. Similarly, in the single-name low frequency trials, while no individual participant saw the same picture in more than one condition, across all participants, a given low frequency single-name picture appeared in the primed condition about half the time and in the unprimed condition about half the time.

We then added the same set of 120 single-name high frequency filler items to each list, and randomized the order of conditions for every individual list, with the restriction that there

were always two intervening trials between primes and targets: that is, (1) between the experimenter's production of a preferred or dispreferred name and the participant's naming of the corresponding dual-name picture or (2) between the experimenter's production of a single low frequency name and the participant's naming of the corresponding picture. In each list, conditions were equally distributed across 5 blocks. To ensure that the participant and experimenter described the same proportion of unprimed pictures (i.e. the first time the picture is presented), and primed pictures (i.e. the second time that the picture is presented), we presented 30 of the high frequency filler trials twice such that they were named first by the participant (production trial) and then by the experimenter (with two intervening trials); the other thirty were named twice by the participant (with two intervening trials). The distribution of item types in each list is shown in Table A2.

	Novel Picture Trial ^a			<i>Mismatch %</i>
	Non-repeated	Repeated after 2 intervening trials:	Repeated Picture Trial ^a	
Production Trial	• 40 high-frequency fillers (20 mismatch) • 10 low-frequency controls (0 mismatch)	• 30 high-frequency fillers (0 mismatch)	• 20 alignment targets (10 primed with preferred name, 10 primed with dispreferred) • 10 low-frequency controls (25% mismatch overall)	25%
Comprehension Trial	• 50 high-frequency fillers (20 mismatch)	• 20 alignment primes (10 preferred, 10 dispreferred; 0 mismatch) • 10 low-frequency controls (0 mismatch)	• 30 high-frequency fillers (25% mismatch overall)	25%

Table A2. Distribution of trials across conditions.

5. Scoring and Extraction of Naming Times in the Collaborative Picture Naming Game

Each participant's entire session was audio-recorded and independently scored by two trained individuals, with disagreements adjudicated by a third trained researcher. The scorers first extracted the names that were produced by participants in each trial. For the dual-name trials, the name produced was automatically marked as "preferred", "dispreferred", or "other", which allowed for creation of one of our key dependent variables: Alignment. Alignment is a binary variable, denoting whether the name produced matched the name used previously by the experimenter (i.e. preferred-preferred or dispreferred-dispreferred).

Additionally, the scorers judged whether each trial was usable for analyses of speech onset latency (naming times). Usable trials were those in which the first sound produced by the speaker was the first phoneme of the name they produced. Responses were excluded from the naming time analyses if (a) they were preceded by a disfluency such as “um” or “uh”, (b) the speaker started to say one name, but then switched to another (e.g. “b-rabbit”), (c) the speaker did not produce a name at all, or (d) there was overlap in speech, e.g. where the experimenter and participant began talking at the same time. Altogether, 17.9% of the *production* trials were excluded (controls: 14.9%; schizophrenia group: 19.4%; bipolar group: 19.4%).

Speech onset times for the produced names were extracted using the Textgrid (silences) function in Praat (Boersma & Weenink, 2016). This processing effectively split the audio recording into segments of silence and sound. Sound segments were manually annotated by three trained scorers who had achieved high inter-rater reliability. From these annotated files, the names produced in the production trials were identified and the speech onset latency was calculated by subtracting the onset time of the picture presentation from the naming onset time. Naming onset times were indexed as the latency between onset of the picture stimulus and the beginning of the participant’s speech.

6. Statistical Analyses

6.1 Modeling approach

Analyses were carried out using lme4 version 1.1-21 (Bates et al., 2015) and lmerTest version 3.1-0 (Kuznetsova et al., 2015) in R (R Core Team, 2022). For linear regression models, we assessed statistical significance using a type-III sums of squares estimation, with p-values

estimated using the Satterthwaite approximation (Satterthwaite, 1946). For logistic regression models, we computed Wald's z .

In cases of model non-convergence, we followed a three-step procedure. First, we allowed the optimization algorithm to run more iterations (up to 100 million). If the model still did not converge, we then tried different optimizers (either the 'bobyqa' or 'Nelder_Mead' optimizer provided by lme4; Bates et al., 2015). If the model still did not converge, we set the correlations between random effects to zero. All models converged after one of these three steps.

6.2 Null Hypothesis Testing

To explore the evidence in favor of the null hypothesis, we performed non-inferiority tests—a one-sided version of the equivalence test (Lakens et al., 2018). This showed that the log odds of alignment in the schizophrenia group and in the bipolar group were non-inferior to the log odds of alignment in controls, given a margin of .386 (equivalent to half a standard deviation from the mean alignment probability in controls).

To provide further evidence for non-inferiority of the schizophrenia group, we calculated Bayes Factors. For the prior, we used a reasonable approximation of an alternative hypothesis predicting inferior alignment in patients (assumption: log odds ratio of alignment in patients to alignment in controls is normally distributed, with a mean of -.5 and a standard deviation of .5) versus the null hypothesis (assumption: difference in log odds of alignment between controls and patients equal to 0). The Bayes Factor was 4.06, such that the evidence for the null hypothesis was approximately four times stronger than evidence for the alternative hypothesis (regardless of which of the two priors we chose). Using Lee & Wagenmaker's (2014) classification, this

provides “moderate” evidence for the null hypothesis. Similar results (Bayes Factor = 4.03) were obtained using a wide, non-informative prior.

6.3 Use of Covariates in Regression Analyses

In certain situations, multicollinearity can influence Type I error rate as a result of inflation of the standard error. This will occur if parameter estimates in the model are unstable (Breagh, 2008; Kalnins, 2018). In such cases, the addition or deletion of a single datapoint or variable can change the results substantially, and one is more likely to get Type I errors for a predictor of interest. Fortunately, some common sense strategies can help detect when significant effects are not real and simply due to the influence of multicollinearity. First, often, but not always, Type I error is associated with a “sign flip” - that is, the sign of the predictor’s parameter estimate, when covariates are added, is opposite its parameter estimate when covariates are excluded (Kalnins, 2018). Other times, the predictor’s parameter estimate may simply become abnormally large when covariates are included in the model. To guard against this, Becker (2005), Breagh (2008) and Kalnins (2018) recommend reporting all bivariate correlations, both amongst predictors and between predictors and the dependent variable. This allows the research to easily detect sign flips and inconsistencies in the parameter estimates. For example, if our dependent variable and predictor of interest are barely correlated (say, $R = .01$) and our parameter estimate in the analysis with covariates is very large, this is likely to be (though not guaranteed to be) a false positive. Similarly, if our predictor and the dependent variable are correlated positively (say, $R = .5$) but the parameter estimate in the analysis with covariates is negative, this is also likely a false positive. In contrast, if the parameter estimates make sense in

light of the zero-order correlations between the predictors and the dependent variable, one can be fairly confident that a Type I error has been avoided.

Relatedly, it may also be useful to report results with and without covariates (Becker, 2005), to show that adding the covariates to the model does not substantially change the results. For example, if the model dependent variable $\sim$ Group produces an effect with p-value of .98 and the model dependent variable $\sim$ Group + Covariates produces an effect with p-value of .001, this is clearly suspicious. Obviously, we expect some minor changes in the model, but reliable parameter estimates for the predictor of interest should generally have the same signs and similar magnitudes, with and without covariates.

In line with the above recommendations, we report bivariate correlations, as well as the results of the regression models without covariates. These additional analyses show that the directionality and magnitude of our parameter estimates make sense in light of the bivariate correlations, that all correlations amongst predictors are mild to moderate (no severe overlap between predictors in any given analysis), and that the parameter estimates are stable (i.e. they don't change much when the covariates are removed).

6.4 Full Regression Results and Model Specifications

Below, we give full results (including nuisance variables) for all between-groups regression analyses reported in the main manuscript. We also give R-style formulas for each model to clarify the model specification.

We also report the results for models with nuisance variables dropped, as well as bivariate correlations between predictors (see above for details on why this was included).

Effect of Group on Log Odds of Lexical Alignment

Alignment ~ Group + Verbal IQ + SES + Age + (1 |subject) + (1 + Group|item)

	Estimate (log odds)	SE	Wald's z	p	Sig.
Intercept	1.02	0.17	6.12	0.00	***
Group (Scz)	0.09	0.19	0.47	0.64	
Group (BP)	0.34	0.18	1.94	0.05	.
Verbal IQ	0.14	0.08	1.81	0.07	.
SES	0.01	0.08	0.19	0.85	
Age	0.03	0.08	0.36	0.72	

Table A3. Logistic mixed effects regression. Statistically significant predictors of interest are shown in bold.

Alignment ~ Group + (1 |subject) + (1 + Group|item)

	Estimate (log odds)	SE	Wald's z	p	Sig.
Intercept	1.08	.17	6.4	.00	***
Group (Scz)	-.05	.17	-.27	.79	
Group (BP)	.32	.18	1.796	.07	.

Table A4. Logistic mixed effects regression—no “nuisance” covariates included in the model. Statistically significant predictors of interest are shown in bold.

	Group: Schizophrenia	Group: Bipolar	Verbal IQ	SES	Age	Alignment
Group:Schizophrenia	1.00	-	-	-	-	-
Group: Bipolar	-0.51	1.00	-	-	-	-
Verbal IQ	-0.41	0.17	1.00	-	-	-
SES	-0.35	0.30	0.38	1.00	-	-
Age	0.32	-0.30	-0.02	-0.28	1.00	-
Alignment	-0.04	0.06	0.06	0.04	0.00	1.00

Table A5. Bivariate correlations between independent and dependent variables. Pearson's r.

Effect of Group on Mentalizing (TASIT Score)

TASIT ~ Group + Verbal IQ + SES + Age

	Estimate	SE	t	p	Sig.
Intercept	53.56	1.06	50.67	0.00	***
Group (Scz) ^a	-3.57	1.62	-2.20	0.03	*
Group (BP) ^b	-2.59	1.48	-1.75	0.08	.
Verbal IQ ^c	4.29	0.69	6.23	0.00	***
SES ^d	0.53	0.68	0.77	0.44	
Age	-1.69	0.65	-2.59	0.01	*

Table A6. Linear regression. Statistically significant predictors of interest are shown in bold.

TASIT ~ Group

	Estimate	SE	t	p	Sig.
Intercept	55.16	1.27	43.38	0.00	***
Group (Scz) ^a	-8.66	1.78	-4.85	0.00	*
Group (BP) ^b	-2.22	1.78	-1.25	0.22	

Table A7. Linear regression—no “nuisance” covariates included in the model. Statistically significant predictors of interest are shown in bold.

	Group: Schizophrenia	Group: Bipolar	Verbal IQ	SES	Age	TASIT
Group:Schizophrenia	1.00	-	-	-	-	-
Group: Bipolar	-0.50	1.00	-	-	-	-
Verbal IQ	-0.42	0.17	1.00	-	-	-
SES	-0.36	0.29	0.38	1.00	-	-
Age	0.30	-0.31	-0.01	-0.27	1.00	-
TASIT	-0.45	0.12	0.64	0.37	-0.25	1.00

Table A8. Bivariate correlations between independent and dependent variables. Pearson’s r.

Effect of Group and Repetition on Naming Response Times

*Log_Naming_RT ~ Repetition*Group + Repetition*Age_Z + Repetition*SES + Repetition*Verbal_IQ + Repetition*Mean_RT + (1 + Repetition | subject) + (1 + Repetition + Group | item)*

	Estimate	SE	t	p	Sig.
Intercept	0.04	0.02	1.72	0.09	.
Repetition	-0.07	0.03	-2.57	0.01	*
Group (Scz)	0.04	0.02	1.53	0.13	
Group (BP)	0.00	0.02	0.14	0.89	
Age	0.00	0.01	0.50	0.62	
SES	0.00	0.01	-0.20	0.84	
Verbal IQ	0.02	0.01	1.96	0.05	.
Mean RT	1.01	0.05	19.93	0.00	***
Group (Scz)*Repetition	-0.07	0.03	-2.39	0.02	*
Group (BP)*Repetition	-0.01	0.03	-0.30	0.76	
Repetition*Age	-0.01	0.01	-0.49	0.63	
Repetition*SES	0.00	0.01	0.35	0.73	
Repetition*Verbal IQ	-0.04	0.01	-2.92	0.00	**
Repetition*Mean RT	0.00	0.07	-0.03	0.97	

Table A9. Mixed effects linear regression. Statistically significant predictors of interest are shown in bold.

*Log_Naming_RT ~ Repetition*Group + (1 + Repetition | subject) + (1 + Repetition + Group | item)*

	Estimate	SE	t	p	Sig.
Intercept	0.30	0.03	8.78	0.00	***
Repetition	-0.08	0.02	-3.73	0.00	***
Group (Scz)	0.30	0.05	6.47	0.00	***
Group (BP)	0.10	0.04	2.16	0.03	*
Group (Scz)*Repetition	-0.05	0.03	-1.83	0.07	.
Group (BP)*Repetition	0.00	0.03	-0.14	0.89	

Table A10. Linear mixed effects regression—no “nuisance” covariates included in the model. Statistically significant predictors of interest are shown in bold.

	Group: Schizophrenia	Group: Bipolar	Verbal IQ	SES	Age	Repetition
Group:Schizophrenia	1.00	-	-	-	-	-
Group: Bipolar	-0.49	1.00	-	-	-	-
Verbal IQ	-0.39	0.15	1.00	-	-	-
SES	-0.34	0.30	0.38	1.00	-	-
Age	0.32	-0.30	0.00	-0.30	1.00	-
Repetition	0.04	-0.02	-0.03	-0.02	0.02	1.00

Table A11. Bivariate correlations between independent and dependent variables. Pearson’s r.

Effect of Group and Alignment on Naming Response Times

Model Specification: $\text{Log_Naming_RT} \sim \text{Alignment} * \text{Group} + \text{Alignment} * \text{Verbal_IQ} + \text{Alignment} * \text{SES} + \text{Alignment} * \text{Age} + (1 + \text{Alignment} | \text{subject}) + (1 + \text{Alignment} * \text{Group} | \text{item})$

	Estimate	SE	t	p	Sig.
Intercept	0.31	0.04	9.07	0.00	***
Alignment	0.02	0.04	-0.20	0.64	
Group (Scz)	0.20	0.07	3.01	0.00	**
Group (BP)	0.12	0.06	2.04	0.05	.
Verbal IQ	-0.04	0.03	-1.76	0.16	
SES	0.02	0.03	0.62	0.44	
Age	0.07	0.03	3.18	0.01	**
Alignment*Group (Scz)	0.02	0.07	2.52	0.81	
Alignment*Group (BP)	-0.07	0.05	-1.01	0.21	
Alignment*Verbal IQ	-0.02	0.03	-0.66	0.33	
Alignment*SES	0.00	0.02	0.10	0.99	
Alignment*Age	-0.01	0.02	-0.62	0.57	

Table A12. Within trials preceded by the dispreferred name; mixed effects linear regression. No predictors of interest were statistically significant.

Model Specification: $\text{Log_Naming_RT} \sim \text{Alignment} * \text{Group} + (1 + \text{Alignment} | \text{subject}) + (1 + \text{Alignment} * \text{Group} | \text{item})$

	Estimate	SE	t	p	Sig.
Intercept	0.30	0.04	8.42	0.00	***
Alignment	-0.01	0.03	-0.37	0.71	
Group (Scz)	0.24	0.05	4.57	0.00	**
Group (BP)	0.08	0.05	1.60	0.11	
Alignment*Group (Scz)	0.01	0.05	0.24	0.81	
Alignment*Group (BP)	-0.04	0.04	-0.82	0.42	

Table A13. Within trials preceded by the dispreferred name; mixed effects linear regression—no “nuisance” covariates in the model. No predictors of interest were statistically significant.

	Group: Schizophrenia	Group: Bipolar	Verbal IQ	SES	Age	Alignment
Group:Schizophrenia	1.00	-	-	-	-	-
Group: Bipolar	-0.48	1.00	-	-	-	-
Verbal IQ	-0.39	0.13	1.00	-	-	-
SES	-0.31	0.30	0.36	1.00	-	-
Age	0.33	-0.32	0.00	-0.29	1.00	-
Alignment	-0.05	0.08	0.12	0.07	-0.02	1.00

Table A14. Bivariate correlations between independent and dependent variables. Pearson's r .

7. Exploratory Analysis: The Relationship Between Alignment and Symptoms

Below, we report the results of an exploratory analysis conducted to examine relationships between lexical alignment and specific relevant symptom clusters in patients with schizophrenia ($n = 32$). Data processing and analysis procedures were the same as for the main analyses of alignment by group.

To index negative symptoms, we used the total PANSS Negative score. To index thought disorder, we used the scores from the Disorganization item of the PANSS. Finally, to explore different facets of social cognition, we included as predictors two scores from the Multnomah Community Ability Scale (MCAS): Social Effectiveness and Social Interest.

The results of this analysis are reported below, in Table A15. We found a small, marginally significant effect of PANSS Negative Total on alignment probability, such that more severe negative symptoms were associated with decreased alignment. Alignment was not predicted by PANSS Disorganization, by MCAS Social Effectiveness, or MCAS Social Interest.

Effect of Symptoms on Alignment Within Patients

Model Specification: Alignment ~ CPZ_Equivalent + PANSS_Negative + PANSS_Disorganization + MCAS_SocEffectiveness + MCAS_SocInterest + (1 | subject) + (1 + PANSS_Negative + PANSS_Disorganization + MCAS_SocEffectiveness + MCAS_SocInterest | item)

	Estimate	SE	t	p	Sig.
Intercept	1.16	0.13	9.22	0.00	***
CPZ Equivalent	0.07	0.08	0.94	0.35	
PANSS – Negative Subscore	-0.19	0.11	-1.71	0.09	.
PANSS – Disorganization Score	-0.12	0.11	-1.12	0.26	
MCAS – Social Effectiveness	-0.06	0.13	-0.45	0.65	
MCAS – Social Interest	0.06	0.10	0.61	0.54	

Table A15. Exploratory analysis of the effect of symptoms on alignment; mixed effects logistic regression. No predictors of interest were statistically significant.

Appendix B

Lexical predictability in schizophrenia: Evidence for selective deficits in use of global context during language production

1. ChatGPT 4.0 Prompt

As noted in the Methods, we used ChatGPT to output a “picture description” text, which we then used to compute Shuffled Surprisal (Unrelated Picture). The prompt we gave to ChatGPT, which mirrors the language of the TLI instructions, is given below.

“I am carrying out a study in which I am asking participants to describe black-and-white pictures.

The instructions to the participants are as follows:

“I want you to describe the picture to me, as fully as you can. Tell me what you see in the picture, describe what you see in this image, and what you think might be happening.”

I would like you pretend that you are an average participant in this study and you are asked to describe a picture of a boy who is sitting with his elbows on the table and in front of him is a violin and bow. He has his head in his hands and is staring sadly at the violin.

Please provide a transcript of a natural utterance that you might produce when asked to describe this picture. You may elaborate on the kinds of things that might be in this picture, but please constrain what you say to what someone might realistically say about a single black and white picture.

It should sound like you are actually talking - not writing.

The language should sound colloquial without a lot of fancy words - the kind of thing that a

regular person might say.

The utterance should be 255 words long.”

ChatGPT’s response is reproduced below:

“Okay, so there's this black-and-white photo, right? And there's this kid, maybe around 10 or 12 years old, sitting at a wooden table. He's got his elbows on the table and he's leaning forward a bit. He's got short, messy hair, and he's wearing a casual button-up shirt with the sleeves rolled up.

Now, right in front of him on the table is a violin and a bow. The violin looks nice and well taken care of. The bow is just lying there next to it, and you can see the strings on it. The black and white in the picture makes everything stand out.

The kid has his head in his hands, like he's really sad or something. He's staring at the violin with this downhearted look on his face. It's like he's been crying or is about to cry, you know? It's kinda touching to see.

In the background, there's a window with some light coming in. It makes some shadows in the room and adds to the whole sad vibe. You can't really see what's outside, but it might be a cloudy day or something.

So, yeah, it looks like the kid is having a tough time with the violin, or maybe something bad happened related to it. The whole thing feels pretty emotional, you know? Like, it really makes you feel for the kid and what he's going through.”

2. Supplementary Figures

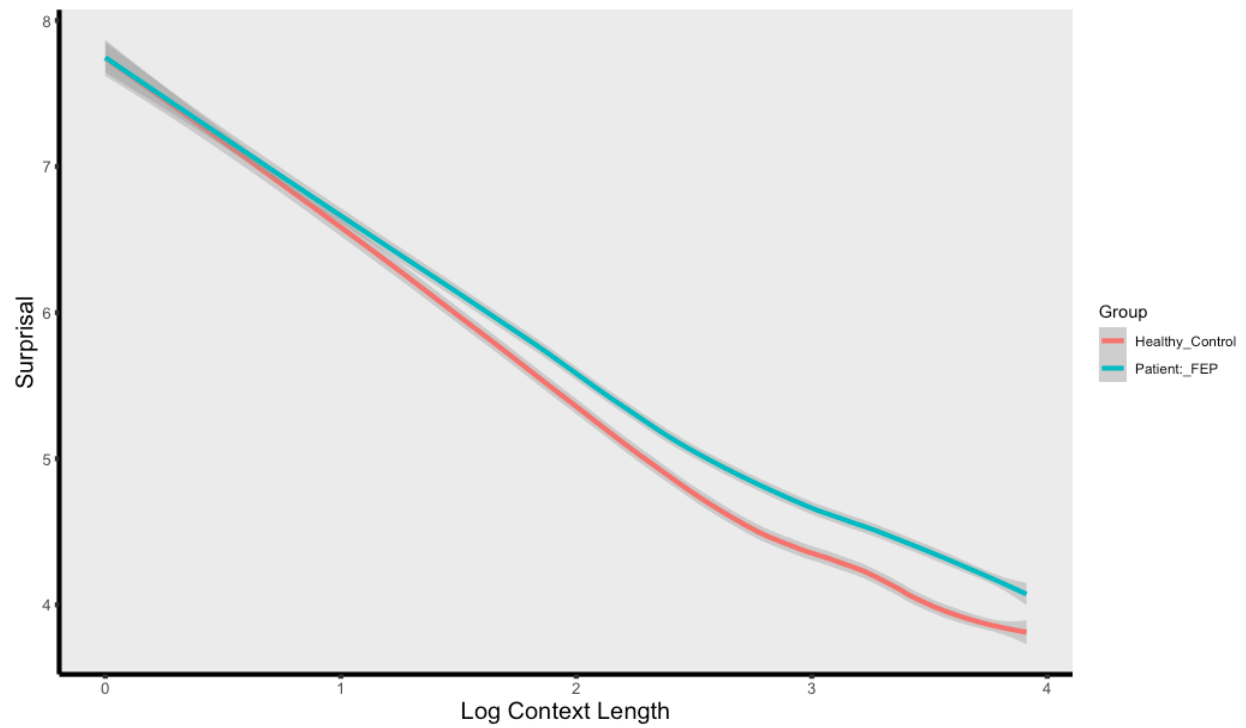


Figure A1. Plot of data by Log Context Length and Group. Lines represent smoothed data using local (“loess”) regression, with span = .5. (Cleveland & Devlin, 1988) Shaded area represents standard error (as estimated by the local regression procedure)

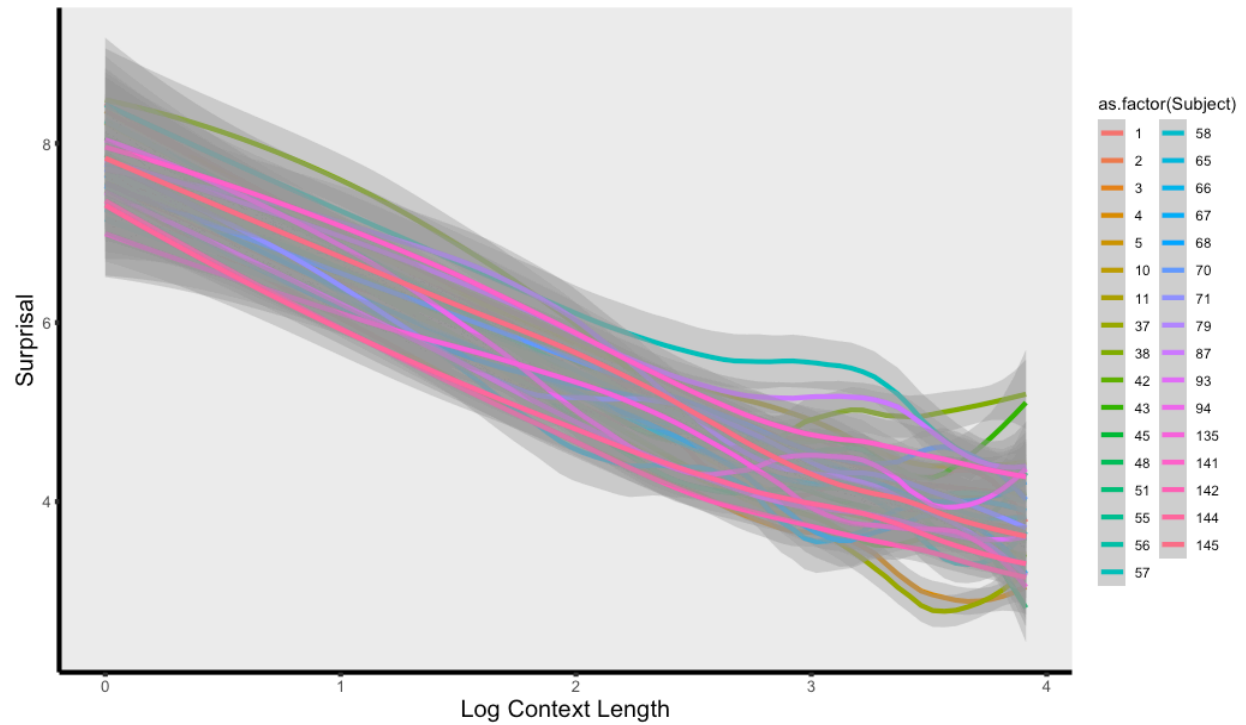


Figure A2. Robustness of the negative effect of Context Length on Surprisal within healthy controls. By-subject lines represent smoothed data using local (“loess”) regression, with span = .5. (Cleveland & Devlin, 1988) Shaded area represents a 95% CI, as estimated by the local regression procedure.

References

- 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
- Allen, H. A. (1983). Do positive symptom and negative symptom subtypes of schizophrenia show qualitative differences in language production? *Psychological Medicine, 13*(4), 787-797. Retrieved from <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0006665095>
- Allen, M. L., Haywood, S., Rajendran, G., & Branigan, H. (2011). Evidence for syntactic alignment in children with autism. *Dev Sci, 14*(3), 540-548.
- American Psychiatric Association. (2013). Schizophrenia Spectrum and Other Psychotic Disorders. In *DSM-V: Diagnostic and Statistical Manual of Mental Disorders*. Arlington, VA: American Psychiatric Publishing.
- 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. Retrieved from http://archpsyc.jamanetwork.com/data/Journals/PSYCH/12320/archpsyc_36_12_006.pdf
- Andreasen, N. C. (1986). Scale for assessment of thought, language and communication (TLC). *Schizophrenia Bulletin, 12*, 473-482. Retrieved from <http://schizophreniabulletin.oxfordjournals.org/content/12/3/473.full.pdf>
- Andreasen, N. C., & Black, D. W. (2005). *Introductory Textbook of Psychiatry*. Washington DC: American Psychiatric Association.

- Andreasen, N. C., & Grove, W. M. (1986). Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophrenia Bulletin*, 12(3), 348-359.
doi:10.1093/schbul/12.3.348
- 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. Retrieved from http://ac.els-cdn.com/0006322393902356/1-s2.0-0006322393902356-main.pdf?_tid=6f687574-869c-11e2-975c-00000aab0f02&acdnat=1362601901_9403645d8dd72ba0a93c779044f8c3d4
- 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
- Arnott, W., Sali, L., & Copland, D. (2011). Impaired reading comprehension in schizophrenia: evidence for underlying phonological processing deficits. *Psychiatry Research*, 187(1-2), 6-10. doi:10.1016/j.psychres.2010.11.025
- Bagner, D. M., Melinder, M. R., & Barch, D. M. (2003). Language comprehension and working memory deficits in patients with schizophrenia. *Schizophrenia Research*, 60(2-3), 299-309. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12591591
- Bambini, V., Arcara, G., Bechi, M., Buonocore, M., Cavallaro, R., & Bosia, M. (2016). The communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Comprehensive Psychiatry*, 71, 106-120.

- Barch, D. M., & Berenbaum, H. (1997). The effect of language production manipulations on negative thought disorder and discourse coherence disturbances in schizophrenia. *Psychiatry Research*, 71, 115-127.
- 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
- Barch, D. M., Cohen, J. D., Servan-Schreiber, D., Steingard, S., Steinhauer, S., & van Kammen, D. (1996). Semantic priming in schizophrenia: An examination of spreading activation using word pronunciation and multiple SOAs. *Journal of Abnormal Psychology*, 105(4), 592-601.
- Barch, D., & Berenbaum, H. (1994). The relationship between information processing and language production. *Journal of Abnormal Psychology*, 103, 241-250.
- Barker, S., Barron, N., McFarland, B. H., & Bigelow, D. A. (1994). A community ability scale for chronically mentally ill consumers: Part I. Reliability and validity. *Community Mental Health Journal*, 30(4), 363-383. doi:10.1007/bf02207489
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
- Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01
- Bazin, N., Sarfati, Y., Lefrere, F., Passerieux, C., & Hardy-Bayle, M. C. (2005). Scale for the evaluation of communication disorders in patients with schizophrenia: a validation study. *Schizophrenia Research*, 77(1), 75-84. doi:10.1016/j.schres.2005.01.020

- Becker, C. A. (1980). Semantic context effects in visual word recognition: an analysis of semantic strategies. *Memory and Cognition*, 8(6), 493-512. doi:10.3758/BF03213769
- Becker, T. E. (2005). Potential Problems in the Statistical Control of Variables in Organizational Research: A Qualitative Analysis With Recommendations. *Organizational Research Methods*, 8(3), 274-289. doi:10.1177/1094428105278021
- 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.
- Bell, A. (2008). Language style as audience design. *Language in Society*, 13(2), 145-204.
- Benjamin, T. B., & Watt, N. F. (1969). Psychopathology and semantic interpretation of ambiguous words. *Journal of Abnormal Psychology*, 74(6), 706-714.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state. A functional MRI study. *Journal of Cognitive Neuroscience*, 11(1), 80-95.
- Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist*, 3(2), 129-136.
- Blaney, P. H. (1974). Two studies on the language behavior of schizophrenics. *Journal of Abnormal Psychology*, 83(1), 23-31. doi:10.1037/h0036230
- Bleuler, E. (1911/1950). *Dementia Praecox, or the Group of Schizophrenias* (J. Zinker, Trans.). New York: International Universities Press.
- Blum, N. A., & Freides, D. (1995). Investigating thought disorder in schizophrenia with the lexical decision task. *Schizophrenia Research*, 16(3), 217-224. doi:10.1016/0920-9964(94)00078-M

- Bora, E., & Pantelis, C. (2016). Social cognition in schizophrenia in comparison to bipolar disorder: A meta-analysis. *Schizophr Res*, 175(1-3), 72-78.
- Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: meta-analysis. *Schizophrenia Research*, 109(1-3), 1-9. doi:10.1016/j.schres.2008.12.020
- 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.
- Bowie, C. R., & Harvey, P. D. (2005). Cognition in schizophrenia: Impairments, determinants, and functional importance. *The Psychiatric Clinics of North America*, 28(3), 613-633, 626.
- 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.
- Bowie, C. R., Gupta, M., & Holshausen, K. (2011). Disconnected and underproductive speech in schizophrenia: unique relationships across multiple indicators of social functioning. *Schizophr Res*, 131(1-3), 152-156.
- Boyce, V., & Levy, R. (2023). A-maze of Natural Stories: Comprehension and surprisal in the Maze task. *Glossa Psycholinguistics*, 2(1). doi:10.5070/g6011190
- Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121(1), 41-57.
- Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers.

- Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(11), 1821.
- Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(11), 1821.
- Breier, A., & Berg, P. H. (2003). The psychosis of schizophrenia. *Biological Psychiatry*, 46, 361-364.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482-1493.
- 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
- 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.
- 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
- Brune, M. (2005). "Theory of mind" in schizophrenia: A review of the literature. *Schizophr Bull*, 31(1), 21-42.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 977-990.

- Buchanan, R. W., Keefe, R. S., Umbricht, D., Green, M. F., Laughren, T., & Marder, S. R. (2011). The FDA-NIMH-MATRICES guidelines for clinical trial design of cognitive-enhancing drugs: what do we know 5 years later? *Schizophrenia Bulletin*, 37(6), 1209-1217. doi:10.1093/schbul/sbq038
- Buck, B., & Penn, D. L. (2015). Lexical Characteristics of Emotional Narratives in Schizophrenia: Relationships With Symptoms, Functioning, and Social Cognition. *J Nerv Ment Dis*, 203(9), 702-708.
- Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1), 8-16. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15597383
- 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
- Cecchi, G., Gutierrez, E., & Corcoran, C. (2019). 24.2 Natural Language Processing Studies of Psychosis and Its Risk States. *Schizophrenia Bulletin*, 45(Supplement_2), S128-S128. doi:10.1093/schbul/sbz022.097
- Chaika, E. (1974). A linguist looks at 'schizophrenic' language. *Brain and Language*, 1, 257-276.
- Chapin, K., McCown, J., Vann, L., Kenney, D., & Youssef, I. (1992). Activation and facilitation in the lexicon of schizophrenics. *Schizophrenia Research*, 6, 251-255. doi:10.1016/0920-9964(92)90008-S
- Chapman, L. J., Chapman, J. P., Curran, T. E., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question.

Developmental Review, 14, 159-185.

Chen, E. Y. H., Wilkins, A. J., & McKenna, P. J. (1994). Semantic memory is both impaired and anomalous in schizophrenia. *Psychological Medicine, 24*, 193-202.

Clare, L., McKenna, P. J., Mortimer, A. M., & Baddeley, A. D. (1993). Memory in schizophrenia: What is impaired and what is preserved? *Neuropsychologia, 31*(11), 1225-1241.

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

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in psychology, 9*, 287-299.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behaviour and biology in schizophrenia. *Psychological Review, 99*(1), 45-77.
doi:10.1037/0033-295x.99.1.45

Colle, L., Angeleri, R., Vallana, M., Sacco, K., Bara, B. G., & Bosco, F. M. (2013).

Understanding the communicative impairments in schizophrenia: a preliminary study. *Journal of Communication Disorders, 46*(3), 294-308.

Condray, R., Steinhauer, S. R., Cohen, J. D., van Kammen, D. P., & Kasperek, A. (1999).

Modulation of language processing in schizophrenia: Effects of context and haloperidol on the event-related potential. *Biological Psychiatry, 45*(10), 1336-1355.

Condray, R., Steinhauer, S. R., van Kammen, D. P., & Kasperek, A. (1996). Working memory capacity predicts language comprehension in schizophrenic patients. *Schizophrenia Research, 20*(1-2), 1-13. Retrieved from <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0008794488>

- 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.
- 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
- Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research, 17*(1), 5-13.
- Crow, T. J. (1997). Schizophrenia as failure of hemispheric dominance for language. *Trends in Neurological Sciences, 20*, 339-343.
- Curtis, V. A., Bullmore, E. T., Brammer, M. J., Wright, I. C., Williams, S. C., Morris, R. G., . . . McGuire, P. K. (1998). Attenuated frontal activation during a verbal fluency task in patients with schizophrenia. *American Journal of Psychiatry, 155*(8), 1056-1063.
- de Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515-535.
doi:10.1353/lan.2006.0130
- Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition, 187*(June 2019), 10-20. doi:10.1016/j.cognition.2019.01.001
- 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

- 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
- Docherty, N. M. (2005). Cognitive impairments and disordered speech in schizophrenia: Thought disorder, disorganization, and communication failure perspectives. *Journal of Abnormal Psychology*, 114(2), 269-278. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15869357
- Docherty, N. M., & Gordinier, S. W. (1999). Immediate memory, attention and communication disturbances in schizophrenia patients and their relatives. *Psychological Medicine*, 29(1), 189-197.
- Doniger, G. M., Silipo, G., Rabinowicz, E. F., Snodgrass, J. G., & Javitt, D. C. (2001). Impaired sensory processing as a basis for object-recognition deficits in schizophrenia. *Am J Psychiatry*, 158(11), 1818-1826.
- Dwyer, K., David, A. S., McCarthy, R., McKenna, P., & Peters, E. (2020). Linguistic alignment and theory of mind impairments in schizophrenia patients' dialogic interactions. *Psychol Med*, 50(13), 2194-2202.
- Eisape, T., Zaslavsky, N., & Levy, R. (2020). Cloze distillation: Improving neural language models with human next-word prediction. In *Proceedings of the 24th Conference on Computational Natural Language Learning* (pp. 609-619). Online: Association for Computational Linguistics.

- 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.
- Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, 36(3), 211-212. doi:10.1017/S0140525X12002312
- First, M. B. (2014). Structured clinical interview for the DSM (SCID). In *The encyclopedia of clinical psychology* (pp. 1-6).
- First, M., Spitzer, R., Miriam, G., & Williams, J. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute.
- Fitzsimmons, J., Kubicki, M., & Shenton, M. E. (2013). Review of functional and anatomical brain connectivity findings in schizophrenia. *Current Opinion in Psychiatry*, 26(2), 172-187. doi:10.1097/YCO.0b013e32835d9e6a
- Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9), 700-711.
- Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in Cognitive Sciences*, 13(4), 175-181. doi:10.1016/j.tics.2009.01.001
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol Rev*, 91(4), 1357-1392. doi:10.1152/physrev.00006.2011
- Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5), 262-268. doi:10.1016/j.tics.2012.04.001

- Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current opinion in neurobiology*, 23(2), 250-254. Retrieved from <http://europepmc.org/abstract/MED/23146876>
- Friston, K. J. (1998). The disconnection hypothesis. *Schizophrenia Research*, 30(2), 115-125. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9549774>
- 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
- Frith, C. D. (2004). Schizophrenia and theory of mind. *Psychological medicine*, 34(03), 385-389.
- Frith, C. D., & Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychol Med*, 26(3), 521-530.
- Frith, C., Friston, K. J., Herold, G., Silberweig, D., Fleicher, P., Cahill, C., . . . Liddle, P. F. (1995). Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task. *British Journal of Psychiatry*, 167, 343-349.
- Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychol Sci*, 23(8), 931-939.
- Garrity, A. G., Pearlson, G. D., McKiernan, K., Lloyd, D., Kiehl, K. A., & Calhoun, V. D. (2007). Aberrant "default mode" functional connectivity in schizophrenia. *American Journal of Psychiatry*, 164(3), 450-457. doi:10.1176/appi.ajp.164.3.450
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181-218.

- Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Geukes, S., Huster, R. J., Wollbrink, A., Junghöfer, M., Zwitserlood, P., & Dobel, C. (2013). A large N400 but no BOLD effect—comparing source activations of semantic priming in simultaneous EEG-fMRI. *PloS One*, 8(12), e84029.
- Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1-68). Cambridge, UK: Cambridge University Press.
- Gold, J. M., Randolph, C., Carpenter, C. J., Goldberg, T. E., & Weinberger, D. R. (1992). Forms of memory failure in schizophrenia. *Journal of Abnormal Psychology*, 101(3), 487-494.
- Goldberg, T. E., Dodge, M., Aloia, M., Egan, M. F., & Weinberger, D. R. (2000). Effects of neuroleptic medications on speech disorganization in schizophrenia: Biasing associative networks towards meaning. *Psychological Medicine*, 30(5), 1123-1130.
- Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., . . . Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, 25(3), 369-380. doi:10.1038/s41593-022-01026-4
- Goodkind, A., & Bicknell, K. (2018). *Predictive power of word surprisal for reading times is a linear function of language model quality*. Paper presented at the Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018), Salt Lake City, Utah, USA.
- Goren, A. R., Tucker, G., & Ginsberg, G. M. (1996). Language dysfunction in schizophrenia. *International Journal of Language and Communication Disorders*, 31(2), 153-170.

- Green, M. F., & Nuechterlein, K. H. (2004). The MATRICS initiative: developing a consensus cognitive battery for clinical trials. *Schizophrenia Research*, 72(1), 1-3.
doi:10.1016/j.schres.2004.09.006
- Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. *Nat Rev Neurosci*, 16(10), 620-631. doi:10.1038/nrn4005
- Green, M. F., Horan, W. P., & Lee, J. (2019). Nonsocial and social cognition in schizophrenia: current evidence and future directions. *World Psychiatry*, 18(2), 146-161.
- Green, M. F., Kern, R. S., Braff, D. L., & Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? *Schizophrenia Bulletin*, 26(1), 119-136. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10755673>
- Grice, P. (1975). *Logic and conversation* (Vol. 3). New York: Academic Press.
- Griego, J. A., Cortes, C. R., Nune, S., Fisher, J. E., & Tagamets, M. A. (2008). Word and letter string processing networks in schizophrenia: evidence for anomalies and compensation. *Brain and Language*, 107(2), 158-166. doi:10.1016/j.bandl.2008.04.001
- Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, 366(6461), 55-58.
- Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annu Rev Neurosci*, 37, 347-362. doi:10.1146/annurev-neuro-071013-013847
- Han, S. D., Nestor, P. G., Hale-Spencer, M., Cohen, A., Niznikiewicz, M., McCarley, R. W., & Wible, C. G. (2007). Functional neuroimaging of word priming in males with chronic schizophrenia. *Neuroimage*, 35(1), 273-282. doi:10.1016/j.neuroimage.2006.11.029

- Hao, Y., Mendelsohn, S., Sterneck, R., Martinez, R., & Frank, R. (2020). Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. *arXiv preprint, arXiv:2009.03954*.
- Harrow, M., Lanin-Kettering, I., & Miller, J. G. (1989). Impaired perspective and thought pathology in schizophrenic and psychotic disorders. *Schizophrenia Bulletin*, 15(4), 605-623.
- Harrow, M., Lanin-Kettering, I., & Miller, J. G. (1989). Impaired perspective and thought pathology in schizophrenic and psychotic disorders. *Schizophrenia Bulletin*, 15(4), 605-623.
- Harrow, M., Tucker, G. J., & Adler, D. (1972). Concrete and idiosyncratic thinking in acute schizophrenic patients. *Archives of General Psychiatry*, 26(5), 433-439.
- Harvey, P. D., Earle-Boyer, E. A., & Wielgus, M. S. (1984). The consistency of thought disorder in mania and schizophrenia. An assessment of acute psychotics. *Journal of Nervous and Mental Disease*, 172(8), 458-463.
- Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *The Journal of Neuroscience*, 28(10), 2539-2550.
- Hayes, R. L., & O'Grady, B. M. (2003). Do people with schizophrenia comprehend what they read? *Schizophrenia Bulletin*, 29(3), 499-507.
- Heilbron, M., Ehinger, B., Hagoort, P., & Lange, F. P. d. (2019). Tracking naturalistic linguistic predictions with deep neural language models. *arXiv preprint, arXiv:1909.04400*.
- Henik, A., Nissimov, E., Priel, B., & Umansky, R. (1995). Effects of cognitive load on semantic priming in patients with schizophrenia. *Journal of Abnormal Psychology*, 104, 576-584. Retrieved from <http://psycnet.apa.org/journals/abn/104/4/576/>

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception.

Trends in Cognitive Sciences, 4(4), 131-138. Retrieved from

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10740277

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature*

Reviews Neuroscience, 8(5), 393-402. doi:10.1038/nrn2113

Hinkley, L. B., Owen, J. P., Fisher, M., Findlay, A. M., Vinogradov, S., & Nagarajan, S. S.

(2010). Cognitive impairments in schizophrenia as assessed through activation and connectivity measures of magnetoencephalography (MEG) data. *Frontiers in Human Neuroscience*, 3, 73. doi:10.3389/neuro.09.073.2009

Hoffman, R. E. (1986). Tree structures, the work of listening, and schizophrenic discourse: A reply to Beveridge and Brown. *Brain and Language*, 27(2), 385-392.

Hoffman, R. E., Kirstein, L., Stopek, S., & Cicchetti, D. V. (1982). Apprehending schizophrenic discourse: A structural analysis of the listener's task. *Brain and Language*, 15(2), 207-233.

Hoffman, R. E., Kirstein, L., Stopek, S., & Cicchetti, D. V. (1982). Apprehending schizophrenic discourse: A structural analysis of the listener's task. *Brain and Language*, 15(2), 207-233.

Hogarty, G. E., & Flesher, S. (1999). Practice principles of cognitive enhancement therapy for schizophrenia. *Schizophrenia Bulletin*, 25(4), 693-708.

Hollingshead, A. B. (1965). *Two factor index of social position*. New Haven, CT: Yale University Press.

- Holzman, P. S., Shenton, M. E., & Solovay, M. R. (1986). Quality of thought disorder in differential diagnosis. *Schizophrenia Bulletin*, 12, 360-371.
- Honea, R., Crow, T. J., Passingham, D., & Mackay, C. E. (2005). Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *American Journal of Psychiatry*, 162, 2233-2245.
- Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 645-662. doi:10.1037/0278-7393.33.4.645
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*, 167(7), 748-751.
- Iter, D., Yoon, J., & Jurafsky, D. (2018). *Automatic detection of incoherent speech for diagnosing schizophrenia*. Paper presented at the Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, New Orleans, LA.
- Johnson, S. C., Lowery, N., Kohler, C., & Turetsky, B. I. (2005). Global-local visual processing in schizophrenia: evidence for an early visual processing deficit. *Biological Psychiatry*, 58(12), 937-946. doi:10.1016/j.biopsych.2005.04.053
- Johnston, M. H., & Holtzman, P. S. (1979). *Assessing Schizophrenic Thinking*. San Francisco: Jossey-Bass.
- Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. *Strategic Management Journal*, 39(8), 2362-2385. doi:10.1002/smj.2783
- Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale

- (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, 13(2), 261-276.
- Keefe, R. (2004). The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophrenia Research*, 68(2-3), 283-297. doi:10.1016/j.schres.2003.09.011
- Keefe, R. S., Harvey, P. D., Goldberg, T. E., Gold, J. M., Walker, T. M., Kennel, C., & Hawkins, K. (2008). Norms and standardization of the Brief Assessment of Cognition in Schizophrenia (BACS). *Schizophr Res*, 102(1-3), 108-115. doi:10.1016/j.schres.2008.03.024
- Kern, R. S., Green, M. F., Fiske, A. P., Kee, K. S., Lee, J., Sergi, M. J., . . . Nuechterlein, K. H. (2009). Theory of mind deficits for processing counterfactual information in persons with chronic schizophrenia. *Psychological Medicine*, 39(4), 645-654. doi:10.1017/S0033291708003966
- Kerns, J. G. (2007). Verbal communication impairments and cognitive control components in people with schizophrenia. *Journal of Abnormal Psychology*, 116(2), 279-289.
- Kerns, J. G., & Berenbaum, H. (2002). Cognitive impairments associated with formal thought disorder in people with schizophrenia. *Journal of Abnormal Psychology*, 111(2), 211-224.
- Kiang, M., Kutas, M., Light, G. A., & Braff, D. L. (2008). An event-related brain potential study of direct and indirect semantic priming in schizophrenia. *American Journal of Psychiatry*, 165(1), 74-81. doi:10.1176/appi.ajp.2007.07050763
- Kircher, T. T., Bullmore, E. T., Brammer, M. J., Williams, S. C., Broome, M. R., Murray, R. M., & McGuire, P. K. (2001). Differential activation of temporal cortex during sentence completion in schizophrenic patients with and without formal thought disorder. *Schizophrenia Research*, 50(1-2), 27-40. Retrieved from

- Kircher, T. T., Liddle, P. F., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2001). Neural correlates of formal thought disorder in schizophrenia: Preliminary findings from a functional magnetic resonance imaging study. *Archives of General Psychiatry*, 58(8), 769-774.
- Kircher, T., Krug, A., Stratmann, M., Ghazi, S., Schales, C., Frauenheim, M., . . . Nagels, A. (2014). A rating scale for the assessment of objective and subjective formal Thought and Language Disorder (TALD). *Schizophrenia Research*, 160(1-3), 216-221.
doi:10.1016/j.schres.2014.10.024
- Klingbeil, J., Wawryzniak, M., Stockert, A., & Saur, D. (2019). Resting-state functional connectivity: An emerging method for the study of language networks in post-stroke aphasia. *Brain and Cognition*, 131, 22-33. doi:j.bandc.2017.08.005
- Koyama, S., Hokama, H., Miyatani, M., Ogura, C., Nageishi, Y., & Shimokochi, M. (1994). ERPs in schizophrenic patients during word recognition task and reaction times. *Electroencephalography and Clinical Neurophysiology*, 92, 546-554.
- Kreher, D. A., Goff, D., & Kuperberg, G. R. (2009). Why all the confusion? Experimental task explains discrepant semantic priming effects in schizophrenia under "automatic" conditions: Evidence from event-related potentials. *Schizophrenia Research*, 111(1-3), 174-181.
- Kreher, D. A., Holcomb, P. J., Goff, D., & Kuperberg, G. R. (2008). Neural evidence for faster and further automatic spreading activation in schizophrenic thought disorder. *Schizophrenia Bulletin*, 34(3), 473-482.
- Kubicki, M., McCarley, R. W., Nestor, P. G., Huh, T., Kikinis, R., Shenton, M. E., & Wible, C. G. (2003). An fMRI study of semantic processing in men with schizophrenia.

- Neuroimage*, 20(4), 1923-1933. Retrieved from
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14683698
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49. doi:10.1016/j.brainres.2006.12.063
- Kuperberg, G. R. (2010). Language in schizophrenia Part 1: an Introduction. *Language and Linguistics Compass*, 4(8), 576-589.
- Kuperberg, G. R. (2010). Language in schizophrenia Part 2: What can psycholinguistics bring to the study of schizophrenia...and vice versa? *Language and Linguistics Compass*, 4(8), 590-604.
- Kuperberg, G. R. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling Reading Comprehension: Behavioral, Neurobiological, and Genetic Components* (pp. 176-192). Baltimore, MD: Paul Brookes Publishing.
- Kuperberg, G. R. (2021). Tea with milk? A Hierarchical Generative Framework of sequential event comprehension. *Topics in Cognitive Science*, 13(1), 256-298.
doi:10.1111/tops.12518
- 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
- Kuperberg, G. R., Broome, M. R., McGuire, P. K., David, A. S., Eddy, M., Ozawa, F., . . . Fischl, B. (2003). Regionally localized thinning of the cerebral cortex in schizophrenia. *Archives of General Psychiatry*, 60(9), 878-888. doi:10.1001/archpsyc.60.9.878

- Kuperberg, G. R., Deckersbach, T., Holt, D. J., Goff, D., & West, W. C. (2007). Increased temporal and prefrontal activity in response to semantic associations in schizophrenia. *Archives of General Psychiatry*, 64(2), 138-151. doi:10.1001/archpsyc.64.2.138
- Kuperberg, G. R., Delaney-Busch, N., Fanucci, K., & Blackford, T. (2018). Priming Production: Neural evidence for enhanced automatic semantic activity immediately preceding language production in schizophrenia. *NeuroImage: Clinical*, 18, 74-85.
- Kuperberg, G. R., Ditman, T., & Choi Perrachione, A. (2018). When Proactivity Fails: An Electrophysiological Study of Establishing Reference in Schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging*, 3(1), 77-87. doi:10.1016/j.bpsc.2017.09.007
- Kuperberg, G. R., McGuire, P. K., & David, A. (1998). Reduced sensitivity to linguistic context in schizophrenic thought disorder: Evidence from online monitoring for words in linguistically-anomalous sentences. *Journal of Abnormal Psychology*, 107, 423-434.
- 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
- Kuperberg, G. R., Weber, K., Delaney-Busch, N., Ustine, C., Stillerman, B., Hamalainen, M., & Lau, E. (2019). Multimodal neuroimaging evidence for looser lexico-semantic networks in schizophrenia: Evidence from masked indirect semantic priming. *Neuropsychologia*, 124, 337-349. doi:10.1016/j.neuropsychologia.2018.10.024
- Kuperberg, G. R., West, W. C., Lakshmanan, B. M., & Goff, D. (2008). fMRI reveals neuroanatomical dissociations during semantic integration in schizophrenia. *Biological Psychiatry*, 64(5), 407-418. doi:10.1016/j.biopsych.2008.03.018

- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647. doi:10.1146/annurev.psych.093008.131123
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-33. <http://cran.r-project.org/package=lmerTest>.
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269.
- Langdon, R., Davies, M., & Coltheart, M. (2002). Understanding minds and understanding communicated meanings in schizophrenia. *Mind & Language*, 17(1-2), 69-104.
- Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484-502.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920-933. doi:10.1038/nrn2532
- Lechner, S., & Northoff, G. (2023). Prolonged Intrinsic Neural Timescales Dissociate from Phase Coherence in Schizophrenia. *Brain Sciences*, 13(4). doi:10.3390/brainsci13040695
- 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.
- Lee, J., Altshuler, L., Glahn, D. C., Miklowitz, D. J., Ochsner, K., & Green, M. F. (2013). Social and nonsocial cognition in bipolar disorder and schizophrenia: relative levels of

- impairment. *Am J Psychiatry*, 170(3), 334-341.
- Lee, J., Rizzo, S., Altshuler, L., Glahn, D. C., Miklowitz, D. J., Sugar, C. A., . . . Green, M. F. (2017). Deconstructing Bipolar Disorder and Schizophrenia: A cross-diagnostic cluster analysis of cognitive phenotypes. *J Affect Disord*, 209, 71-79.
- Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8), 2906-2915. doi:10.1523/JNEUROSCI.3684-10.2011
- Levine, J., Schild, K., Kimhi, R., & Schreiber, G. (1996). Word associative production in affective versus schizophrenic psychoses. *Psychopathology*, 29(1), 7-13. Retrieved from <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0008711078>
- Li, X., Branch, C. A., Nierenberg, J., & Delisi, L. E. (2010). Disturbed functional connectivity of cortical activation during semantic discrimination in patients with schizophrenia and subjects at genetic high-risk. *Brain Imaging and Behavior*, 4(1), 109-120. doi:10.1007/s11682-010-9090-3
- 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
- Liemburg, E. J., Vercammen, A., Ter Horst, G. J., Curcic-Blake, B., Knegtering, H., & Aleman, A. (2012). Abnormal connectivity between attentional, language and auditory networks in schizophrenia. *Schizophrenia Research*, 135(1-3), 15-22. doi:10.1016/j.schres.2011.12.003

- 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
- Mackinley, M., Chan, J., Ke, H., Dempster, K., & Palaniyappan, L. (2021). Linguistic determinants of formal thought disorder in first episode psychosis. *Early Intervention Psychiatry*, 15(2), 344-351. doi:10.1111/eip.12948
- Maeda, K., Kasai, K., Uetsuki, M., Hata, A., Araki, T., Rogers, M. A., . . . Iwanami, A. (2007). Increased positive thought disorder with illness duration in patients with schizophrenia. *Psychiatry and Clinical Neurosciences*, 61(6), 687-690. Retrieved from <http://dx.doi.org/10.1111/j.1440-1819.2007.01730.x>
- 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.
- Marengo, J. T., & Harrow, M. (1987). Schizophrenic thought disorder at follow-up. *Archives of General Psychiatry*, 44(7), 651-659. doi:10.1001/archpsyc.1987.01800190071011
- Marengo, J. T., & Harrow, M. (1997). Longitudinal course of thought disorder in schizophrenia and schizoaffective disorder. *Schizophrenia Bulletin*, 23, 273-285.
- 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.

- Marengo, J. T., Harrow, M., Lanin-Kettering, I., & Wilson, A. (1986). Evaluating bizarre-
idiosyncratic thinking: A comprehensive index of positive thought disorder.
Schizophrenia Bulletin, 12, 497-509.
- Marques, V. R., Spataro, P., Cestari, V., Sciarretta, A., & Rossi-Arnaud, C. (2016). Testing the
Identification/Production Hypothesis of Implicit Memory in Schizophrenia: The Role of
Response Competition. *Journal of the International Neuropsychological Society*, 22(3),
314-321. doi:10.1017/S1355617715001198
- McDonald, S., Bornhofen, C., Shum, D., Long, E., Saunders, C., & Neulinger, K. (2006).
Reliability and validity of The Awareness of Social Inference Test (TASIT): a clinical
test of social perception. *Disabil Rehabil*, 28(24), 1529-1542.
- McGuire, P. K., Quested, D. J., Spence, S. A., Murray, R. M., Frith, C. D., & Liddle, P. F. (1998).
Pathophysiology of 'positive' thought disorder in schizophrenia. *British Journal of
Psychiatry*, 173, 231-235.
- McGurk, S. R., & Meltzer, H. Y. (2000). The role of cognition in vocational functioning in
schizophrenia. *Schizophrenia Research*, 45, 175-184.
- Menon, R. R., Barta, P. E., Aylward, E. H., Richards, S. S., Vaughn, D. D., Tien, A. Y., . . .
Pearlson, G. D. (1995). Posterior superior temporal gyrus in schizophrenia: Grey matter
changes and clinical correlates. *Schizophrenia Research*, 16(2), 127-135.
- Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., & Coulson, S. (2023).
Strong Prediction: Language Model Surprisal Explains Multiple N400 Effects.
Neurobiology of Language, 1-71. doi:10.1162/nol_a_00105

- Millet, J., Caucheteux, C., Orhan, P., Boubenec, Y., Gramfort, A., Dunbar, E., . . . King, J. (2022). *Toward a realistic model of speech processing in the brain with self-supervised learning*. Paper presented at the Advances in Neural Information Processing Systems.
- 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.
- 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.
- Mitchell, R. L., & Crow, T. J. (2005). Right hemisphere language functions and schizophrenia: The forgotten hemisphere? *Brain*, 128(Pt 5), 963-978. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15743870
- Mitchley, N. J., Barber, J., Gray, J. M., Brooks, D. N., & Livingston, M. G. (1998). Comprehension of irony in Schizophrenia. *Cognitive Neuropsychiatry*, 3(2), 127-138. doi:10.1080/135468098396206
- Moran, L. J., Mefferd, R. B., & Kimble, J. P. (1964). Idiodynamic sets in word association. *Psychological Monographs: General and Applied*, 78, 1-22.
- Moritz, S., Mersmann, K., Kloss, M., Jacobsen, D., Andresen, B., Krausz, M., . . . Naber, D. (2001). Enhanced semantic priming in thought-disordered schizophrenic patients using a word pronunciation task. *Schizophrenia Research*, 48(2-3), 301-305. Retrieved from

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11295382

Moritz, S., Mersmann, K., Kloss, M., Jacobsen, D., Wilke, U., Andresen, B., . . . Pawlik, K.

(2001). 'Hyper-priming' in thought-disordered schizophrenic patients. *Psychological Medicine*, 31(2), 221-229. Retrieved from

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11232910

Mota, N. B., Carrillo, F., Slezak, D. F., Copelli, M., & Ribeiro, S. (2016). *Characterization of the relationship between semantic and structural language features in psychiatric diagnosis*.

Paper presented at the 2016 50th Asilomar Conference on Signals, Systems, and Computers.

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

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

Mueser, K. T., Bellack, A. S., Douglas, M. S., & Morrison, R. L. (1991). Prevalence and stability of social skill deficits in schizophrenia. *Schizophrenia Research*, 5, 167-176.

Nadig, A., Seth, S., & Sasson, M. (2015). Global similarities and multifaceted differences in the production of partner-specific referential pacts by adults with autism spectrum disorders. *Front Psychol*, 6, 1888.

- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106, 226-254.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic Processes in Reading and Visual Word Recognition* (pp. 264-333). Hillsdale, NJ: Erlbaum.
- Neely, J. H., Keefe, D. E., & Ross, K. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1003-1019. doi:10.1037/0278-7393.15.6.1003
- 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
- Nestor, P. G., Akdag, S. J., O'Donnell, B. F., Niznikiewicz, M., Law, S., Shenton, M. E., & McCarley, R. W. (1998). Word recall in schizophrenia: A connectionist model. *American Journal of Psychiatry*, 155(12), 1685-1690.
- Nestor, P. G., Kimble, M. O., O'Donnell, B. F., Smith, L., Niznikiewicz, M., Shenton, M. E., & McCarley, R. W. (1997). Aberrant semantic activation in schizophrenia: A neurophysiological study. *American Journal of Psychiatry*, 154(5), 640-646.
- Niznikiewicz, M. A., O'Donnell, B. F., Nestor, P. G., Smith, L., Law, S., Karapellou, M., . . . McCarley, R. W. (1997). ERP assessment of visual and auditory language processing in schizophrenia. *Journal of Abnormal Psychology*, 106, 85-94.

- Northoff, G., Sandsten, K. E., Nordgaard, J., Kjaer, T. W., & Parnas, J. (2021). The Self and Its Prolonged Intrinsic Neural Timescale in Schizophrenia. *Schizophrenia Bulletin*, 47(1), 170-179. doi:10.1093/schbul/sbaa083
- Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (under review). A predictive coding model of the N400. *bioRxiv*. doi:10.1101/2023.04.10.536279
- Ober, B. A., Vinogradov, S., & Shenaut, G. K. (1995). Semantic priming of category relations in schizophrenia. *Neuropsychology*, 9(2), 220-228. doi:Doi 10.1037//0894-4105.9.2.220
- Oltmanns, T. F., Murphy, R., Berenbaum, H., & Dunlop, S. R. (1985). Rating verbal communication impairment in schizophrenia and affective disorders. *Schizophrenia Bulletin*, 11(2), 292-299. Retrieved from <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0004012229>
- 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
- Palaniyappan, L., Mota, N. B., Oowise, S., Balain, V., Copelli, M., Ribeiro, S., & Liddle, P. F. (2019). Speech structure links the neural and socio-behavioural correlates of psychotic disorders. *Prog Neuropsychopharmacol Biol Psychiatry*, 88, 112-120. doi:10.1016/j.pnpbp.2018.07.007
- Patterson, T. L., Moscona, S., McKibbin, C. L., Davidson, K., & Jeste, D. V. (2001). Social skills performance assessment among older patients with schizophrenia. *Schizophrenia Research*, 48(2), 351-360.
- Pawełczyk, A., Kotlicka-Antczak, M., Łojek, E., Ruszpel, A., & Pawełczyk, T. (2017). Schizophrenia patients have higher-order language and extralinguistic impairments.

Schizophrenia Research.

- Perkins, D. O., & McGlashan, T. H. (2015). Severity of thought disorder predicts psychosis in persons at clinical high-risk. *Schizophrenia Research*. doi:10.1016/j.schres.2015.09.008
- Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behav Brain Sci*, 27(2), 212-225.
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2-3), 203-228.
- Pinkham, A. E., Penn, D. L., Green, M. F., & Harvey, P. D. (2016). Social cognition psychometric evaluation: Results of the initial psychometric study. *Schizophr Bull*, 42(2), 494-504.
- Pomarol-Clotet, E., Oh, T. M., Laws, K. R., & McKenna, P. J. (2008). Semantic priming in schizophrenia: Systematic review and meta-analysis. *British Journal of Psychiatry*, 192(2), 92-97. doi:10.1192/bjp.bp.106.032102
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: the Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2), 816-847. doi:10.1016/j.neuroimage.2012.04.062
- Prigatano, G. P., Pepping M., & Klonoff, P. (1986). *Cognitive, personality, and psychosocial factors in the neuropsychological assessment of brain-injured patients*.

- Qian, T., & Jaeger, T. F. (2010). *Close = Relevant? The role of context in efficient language production*. Paper presented at the Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics, Uppsala, Sweden.
- R Core Team. (2022). R: A language and environment for statistical computing (Version 4.2.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- 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.
- Racenstein, J. M., Penn, D., Harrow, M., & Schleser, R. (1999). Thought disorder and psychosocial functioning in schizophrenia: The concurrent and predictive relationships. *Journal of Nervous and Mental Disease*, 187, 281-289.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9. Retrieved from <https://openai.com/blog/better-language-models/>
- 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.
- 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.

- Ragland, J. D., Gur, R. C., Valdez, J., Turetsky, B. I., Elliott, M., Kohler, C., . . . Gur, R. E. (2004). Event-related fMRI of frontotemporal activity during word encoding and recognition in schizophrenia. *American Journal of Psychiatry*, 161(6), 1004-1015.
- Ragland, J. D., Moelter, S. T., Bhati, M. T., Valdez, J. N., Kohler, C. G., Siegel, S. J., . . . Gur, R. E. (2008). Effect of retrieval effort and switching demand on fMRI activation during semantic word generation in schizophrenia. *Schizophrenia Research*, 99(1-3), 312-323. doi:10.1016/j.schres.2007.11.017
- Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76, 29-46.
- Rimol, L. M., Hartberg, C. B., Nesvag, R., Fennema-Notestine, C., Hagler, D. J., Jr., Pung, C. J., . . . Agartz, I. (2010). Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. 68(1), 41-50. doi:10.1016/j.biopsych.2010.03.036
- Roberts, D. L., & Penn, D. L. (2009). Social cognition and interaction training (SCIT) for outpatients with schizophrenia: a preliminary study. *Psychiatry Research*, 166(2), 141-147.
- Rochester, S., Martin, J. R., & Thurston, S. (1977). Thought-process disorder in schizophrenia: The listener's task. *Brain and Language*, 4(1), 95-114.
- Ruchsow, M., Trippel, N., Groen, G., Spitzer, M., & Kiefer, M. (2003). Semantic and syntactic processes during sentence comprehension in patients with schizophrenia: Evidence from event-related potentials. *Schizophrenia Research*, 64(2-3), 147-156.
- Rutter, D. R. (1985). Language in schizophrenia. The structure of monologues and conversations. *British Journal of Psychiatry*, 146, 399-404.
- S., M., Flanagan, S., Rollins, J., & Kinch, J. (2003). TASIT: A new clinical tool for assessing

- social perception after traumatic brain injury. *J Head Trauma Rehabil*, 18(3), 219-238.
- Salzinger, K., Pisoni, D. B., Portnoy, S., & Feldman, R. S. (1970). The immediacy hypothesis and response-produced stimuli in schizophrenic speech. *Journal of Abnormal Psychology*, 76, 258-264.
- Salzinger, K., Portnoy, S., & Feldman, R. S. (1964). Verbal behavior of schizophrenic and normal subjects. *Annals of the New York Academy of Sciences*, 105, 845-860.
- Samame, C. (2013). Social cognition throughout the three phases of bipolar disorder: A state-of-the-art overview. *Psychiatry Res*, 210(3), 1275-1286.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6), 110-114.
- Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija*, 43(4), 441-464. doi:10.2298/Psi1004441s
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in neurosciences*, 26(2), 100-107.
- Searles, H. F. (1962). The differentiation between concrete and metaphorical thinking in the recovering schizophrenic patient. *Journal of the American Psychoanalytic Association*, 10(1), 22-49.
- Sharpe, V., Weber, K., & Kuperberg, G. R. (2020). Impairments in probabilistic prediction and Bayesian learning can explain reduced neural semantic priming in schizophrenia. *Schizophr Bull*, 46(6), 1558-1566.
- Shenton, M. E., Kikinis, R., Jolesz, F. A., Pollak, S. D., LeMay, M., Wible, C. G., . . . McCarley, R. W. (1992). Abnormalities of the left temporal lobe and thought disorder in schizophrenia. *New England Journal of Medicine*, 327, 604-612.

- Shenton, M. E., Solovay, M. R., Holzman, P. S., Coleman, M., & Gale, H. J. (1989). Thought disorder in the relatives of psychotic patients. *Archives of General Psychiatry*, 46(10), 897-901.
- Sichlinger, L., Cibelli, E., Goldrick, M., & Mittal, V. A. (2019). Clinical correlates of aberrant conversational turn-taking in youth at clinical high-risk for psychosis. *Schizophrenia Research*, 204, 419-420. doi:10.1016/j.schres.2018.08.009
- Silva, A., Limongi, R., MacKinley, M., & Palaniyappan, L. (2021). Small Words That Matter: Linguistic Style and Conceptual Disorganization in Untreated First-Episode Schizophrenia. *Schizophrenia Bulletin Open*, 2(1), sgab010. doi:10.1093/schizbullopen/sgab010
- Sitnikova, T., Salisbury, D. F., Kuperberg, G. R., & Holcomb, P. J. (2002). Electrophysiological insights into language processing in schizophrenia. *Psychophysiology*, 39(6), 851-860. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12462512>
- Slocombe, K. E., Alvarez, I., Branigan, H. P., Jellema, T., Burnett, H. G., Fischer, A., . . . Levita, L. (2013). Linguistic alignment in adults with and without Asperger's syndrome. *J Autism Dev Disord*, 43(6), 1423-1436.
- Smith, N. J., & Levy, R. (2011). *Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing*. Paper presented at the CogSci.
- 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
- Smitha, K. A., Akhil Raja, K., Arun, K. M., Rajesh, P. G., Thomas, B., Kapilamoorthy, T. R., & Kesavadas, C. (2017). Resting state fMRI: A review on methods in resting state

- connectivity analysis and resting state networks. *The Neuroradiology Journal*, 30(4), 305-317. doi:10.1177/1971400917697342
- Soler, M. J., Ruiz, J. C., Fuentes, I., & Tomas, P. (2007). A comparison of implicit memory tests in schizophrenic patients and normal controls. *The Spanish Journal of Psychology*, 10(2), 423-429.
- Solomon, M., Olsen, E., Niendam, T., Ragland, J. D., Yoon, J., Minzenberg, M., & Carter, C. S. (2011). From lumping to splitting and back again: atypical social and language development in individuals with clinical-high-risk for psychosis, first episode schizophrenia, and autism spectrum disorders. *Schizophrenia Research*, 131(1–3), 146-151. doi:10.1016/j.schres.2011.03.005
- 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.
- Sparks, A., McDonald, S., Lino, B., O'Donnell, M., & Green, M. J. (2010). Social cognition, empathy and functional outcome in schizophrenia. *Schizophrenia Research*, 122(1-3), 172-178.
- Spencer, T. J., Thompson, B., Oliver, D., Diederer, 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
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford, UK: Blackwell.

- Spitzer, M. (1997). A cognitive neuroscience view of schizophrenic thought disorder. *Schizophrenia Research*.
- Spitzer, M., Beuckers, J., Beyer, S., Majer, S., & Hermle, L. (1994). Contextual insensitivity in thought-disordered schizophrenic patients: Evidence from pauses in spontaneous speech. *Language and Speech*, 37(2), 171-185.
- Sponheim, S. R., Steele, V. R., & McGuire, K. A. (2004). Verbal memory processes in schizophrenia patients and biological relatives of schizophrenia patients: intact implicit memory, impaired explicit recollection. *Schizophr Res*, 71(2-3), 339-348.
- Sponheim, S. R., Surerus-Johnson, C., Leskela, J., & Dieperink, M. E. (2003). Proverb interpretation in schizophrenia: the significance of symptomatology and cognitive processes. *Schizophr Res*, 65(2-3), 117-123. doi:10.1016/s0920-9964(02)00525-x
- Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing*, 17(3), 279-305. doi:10.1007/s10339-016-0765-6
- Staub, A. (2015). The effect of lexical predictability on eye movements in reading: critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8), 311-327. doi:10.1111/lnc3.12151
- Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, 82, 1-17. doi:10.1016/j.jml.2015.02.004
- 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
- Stewart, S. L., Corcoran, R., & Drake, R. J. (2008). Alignment and theory of mind in

- schizophrenia. *Cognitive neuropsychiatry*, 13(5), 431-448.
- Sun, J., Maller, J. J., Guo, L., & Fitzgerald, P. B. (2009). Superior temporal gyrus volume change in schizophrenia: a review on region of interest volumetric studies. *Brain Research Reviews*, 61(1), 14-32. doi:10.1016/j.brainresrev.2009.03.004
- 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
- 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
- Tan, E. J., & Rossell, S. L. (2014). Building a neurocognitive profile of thought disorder in schizophrenia using a standardized test battery. *Schizophr Res*, 152(1), 242-245. doi:10.1016/j.schres.2013.11.001
- Tan, E. J., & Rossell, S. L. (2019). Language comprehension and neurocognition independently and concurrently contribute to formal thought disorder severity in schizophrenia. *Schizophr Res*, 204, 133-137. doi:10.1016/j.schres.2018.08.019
- Tan, E. J., Neill, E., Tomlinson, K., & Rossell, S. L. (2020). Semantic Memory Impairment Across the Schizophrenia Continuum: A Meta-Analysis of Category Fluency Performance. *Schizophrenia Bulletin Open*, 1(1). doi:10.1093/schizbullopen/sgaa054
- 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.comppsy.2013.10.016

- Taylor, W. (1953). 'Cloze' procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- Thermenos, H. W., Whitfield-Gabrieli, S., Seidman, L. J., Kuperberg, G. R., Juelich, R. J., Divatia, S., . . . DeLisi, L. E. (2013). Altered language network activity in young people at familial high-risk for schizophrenia. *Schizophrenia Research*, 151(1–3), 229-237. doi:10.1016/j.schres.2013.09.023
- Thoma, P., Hennecke, M., Mandok, T., Wahner, A., Brune, M., Juckel, G., & Daum, I. (2009). Proverb comprehension impairments in schizophrenia are related to executive dysfunction. *Psychiatry Research*, 170(2-3), 132-139. doi:10.1016/j.psychres.2009.01.026
- 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.
- Tomasi, D., & Volkow, N. D. (2012). Resting functional connectivity of language networks: characterization and reproducibility. *Molecular Psychiatry*, 17(8), 841-854.
- Tweedy, J. R., & Lapinski, R. H. (1981). Facilitating word recognition: Evidence for strategic and automatic factors. *The Quarterly Journal of Experimental Psychology Section A*, 33(1), 51-59. doi:10.1080/14640748108400768
- Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition*, 5(1), 84-89.
- Uscatescu, L. C., Kronbichler, M., Said-Yurekli, S., Kronbichler, L., Calhoun, V., Corbera, S., . . . Assaf, M. (2023). Intrinsic neural timescales in autism spectrum disorder and

- schizophrenia. A replication and direct comparison study. *Schizophrenia (Heidelb)*, 9(1), 18. doi:10.1038/s41537-023-00344-1
- Uscatescu, L. C., Said-Yurekli, S., Kronbichler, L., Stelzig-Scholer, R., Pearce, B. G., Reich, L. A., . . . Kronbichler, M. (2021). Reduced intrinsic neural timescales in schizophrenia along posterior parietal and occipital areas. *NPJ Schizophrenia*, 7(1), 55. doi:10.1038/s41537-021-00184-x
- Vaskinn, A., & Horan, W. P. (2020). Social Cognition and Schizophrenia: Unresolved Issues and New Challenges in a Maturing Field of Research. *Schizophr Bull*, 46(3), 464-470.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). *Attention is all you need*. Paper presented at the Advances in Neural Information Processing Systems, Long Beach, CA, USA.
- Vinogradov, S., Ober, B. A., & Shenaut, G. K. (1992). Semantic priming of word pronunciation and lexical decision in schizophrenia. *Schizophrenia Research*, 8, 171-181.
- 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*.
- 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
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale, Revised*. New York: Psychological Corporation.

- Wheeler, A. L., & Voineskos, A. N. (2014). A review of structural neuroimaging in schizophrenia: from connectivity to connectomics. *8*, 653.
doi:10.3389/fnhum.2014.00653
- Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., & Levy, R. P. (2020). On the predictive power of neural language models for human real-time comprehension behavior. *arXiv preprint, arXiv:2006.01912*.
- Wilson, D., & Sperber, D. (2002). Truthfulness and relevance. *Mind*, *111*(443), 583-632.
- Wilson, L. B., Rojas, D. C., Shatti, S., & Tregellas, J. R. (2013). Greater neuronal responses during automatic semantic processing in schizophrenia. *Neuroreport*, *24*(5), 212.
doi:10.1097/WNR.0b013e32835eb688
- Wisco, J. J., Kuperberg, G. R., Manoach, D., Quinn, B. T., Busa, E., Fischl, B., . . . Sorensen, A. G. (2007). Abnormal cortical folding patterns within Broca's area in schizophrenia: evidence from structural MRI. *Schizophrenia Research*, *94*(1-3), 317-327.
doi:10.1016/j.schres.2007.03.031
- Wolf, D. H., Gur, R. C., Valdez, J. N., Loughhead, J., Elliott, M. A., Gur, R. E., & Ragland, J. D. (2007). Alterations of fronto-temporal connectivity during word encoding in schizophrenia. *Psychiatry Research*, *154*(3), 221-232.
doi:10.1016/j.psychresns.2006.11.008
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., . . . Rush, A. M. (2020). *Transformers: State-of-the-art natural language processing*. Paper presented at the In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, Brooklyn, USA.

- Woodward, T. S., Tipper, C. M., Leung, A., Lavigne, K. M., Sanford, N., & Metzak, P. D. (2015).
Reduced functional connectivity during controlled semantic integration in schizophrenia:
A multivariate approach. *Human Brain Mapping*. doi:10.1002/hbm.22820
- Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for mania:
reliability, validity and sensitivity. *The British Journal of Psychiatry*, 133(5), 429-435.