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

Table of Contents

1. Terminology

1.1 Discussion of "Mentalizing" Terminology

We note that throughout this manuscript, we use the terms "mentalizing" and "perspective-taking" interchangeably, to mean "the process by which we make sense of mental states in other people." In essence, these terms refer to "putting oneself in another person's shoes," an ability which is crucial for effective communication. We cannot know whether our partner is understanding us unless we try to see things from their perspective.

Notably, these terms are often used synonymously with "theory of mind", though the reader will note that we do not use this term in the present paper. This decision was made in order to avoid some of the assumptions that may come along with this term, as it has been used in different ways throughout the literature. Some describe theory of mind, not as a process or an ability, but as a core assumption with which healthy adults reason about the world: the assumption that behavior is caused by mental states (Premack & Woodruff, 1978; Frith & Frith, 2006). That is, theory of mind is not something we *use* but something we *have*. In contrast, others have argued that we ought to think of "theory of mind" not as a literal theory that develops over time, but as a mechanism by which the concepts of belief and desire are introduced to the cognitive system (Leslie, Friedman, & German, 2004). Similarly, it has become common, particularly in psychopathology research, to talk about theory of mind as an ability that can be more or less impaired, measured on a scale of low to high (e.g. Fretland et al., 2015; Bora, 2017). Each of these uses carries different assumptions about the cognitive architecture, which are beyond the scope of the present paper. Thus, to avoid any confusion, we stick to the terms "mentalizing" and "perspective-taking," which unambiguously refer to the process of inferring another person's mental state.

1.2 Use of the terms "Top-Down" and "Bottom-Up" in relation to Priming and Mechanisms of Alignment

The concepts of "top-down" and "bottom-up" processing are key to the present paper. We use the term "top-down" to refer to the flow of information from *higher levels* of representation (encoding information at a larger spatiotemporal scale) to *lower levels* of representation (encoding information at a smaller spatiotemporal scale). Conversely, we use the term "bottom-up" to refer to the flow of information from *lower* to *higher* levels of representation. Importantly, in a hierarchically organized system (like language), what counts as a "high-level" and what counts as a "low-level" of representation is always relative, depending on the level of representation in which we are interested. In the present study, we are interested in the top-down and bottom-up influences on *lexical* representations (features associated with individual words) during lexical priming and lexical alignment.

In the case of lexical priming, *lower-level* information that might influence the magnitude of the priming effect through *bottom-up* mechanisms might include the degree of orthographic, phonological and semantic overlap between the preceding prime word and the target. *Higher-level* information that might influence the magnitude of the priming effect through *top-down* mechanisms might include goal representations (e.g. those determined by task requirements, see Neely, 1991) for a review) or representations of the predictive validity of the broader contextual environment (e.g. our beliefs about the relative proportions of related versus unrelated prime-target pairs, see Tweedy & Lapinski, 1981; Neely, 1991; Lau, Holcomb & Kuperberg, 2013; Delaney-Busch, Morgan, Lau & Kuperberg, 2019).

Analogously, in the case of lexical alignment, *lower-level* information that might influence the magnitude of the alignment effect through *bottom-up* mechanisms might include the phonological or semantic overlap between the word our conversational partner has previously spoken, and the word that we are about to utter (Pickering & Garrod, 2004). *Higher-level* information that might influence alignment through *top-down* mechanisms might include our communicative goals, including our beliefs that our partner might understand us better if we use the same words as them to refer to objects in our common environment (e.g. using the word “bunny” rather than “rabbit”; Branigan et al., 2011). In the present paper, we refer to this top-down modulation as *mentalizing* or *perspective-taking* (see above section for discussion of this terminology).

2. 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 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. They were completed within eight weeks of the experimental session.

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.

3. The Collaborative 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 that was 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). To narrow these down, we 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.

In the first norming study, 50 participants were presented with 60 photographs: the 45 items that we thought would elicit multiple names, and 15 filler items that we thought would 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 not stated) 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, and 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%)). This norming study resulted in the selection of 32 pictures.

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 rating of 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 the English Lexicon project: http://elexicon.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 Supplementary Table 1.

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

We additionally developed 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 SUBLTEX 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 these 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	Mushroo	Pineapple

		m	
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

Supplementary Table 1. 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. Alignment targets, and low frequency items (matched for frequency with dispreferred names).

<u>*3.3 High frequency single-name pictures: Fillers*</u>

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 (Brysbaert & New, 2009). 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 *nth* 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 Supplementary Table 2.

	Novel Picture Trial[a]			
	Non-repeated	**Repeated after 2 intervening trials:**	**Repeated Picture Trial[a]**	***Mismatch %***
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%*

Supplementary Table 2. 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, 2016). 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 (Breaugh, 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), Breaugh (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 ~ Group produces an effect with p-value of .98

and the model dependent variable ~ 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	

Supplementary Table 3A. 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	.

Supplementary Table 3B. 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

Supplementary Table 3C. 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	*

Supplementary Table 4A. 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	

Supplementary Table 4B. 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

Supplementary Table 4C. 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	

Supplementary Table 5A. 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	

Supplementary Table 5B. 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

Supplementary Table 5C. Bivariate correlations between independent and dependent variables. Pearson's r.

Effect of Group and Alignment on Naming Response Times

*Model Specification: Log_Naming_RT ~ Alignment*Group + Alignment*Verbal_IQ + Alignment*SES + Alignment*Age + (1 +Alignment|subject) + (1 + Alignment*Group |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	

Supplementary Table 6A. Within trials preceded by the dispreferred name; mixed effects linear regression. No predictors of interest were statistically significant.

*Model Specification: Log_Naming_RT ~ Alignment*Group + (1 +Alignment|subject) + (1 + Alignment*Group |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	

Supplementary Table 6B. 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

Supplementary Table 6C. 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 Supplementary Table 7. 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	

Supplementary Table 7. Exploratory analysis of the effect of symptoms on alignment; mixed effects logistic regression. No predictors of interest were statistically significant.

8. References

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.

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.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (version 6.0. 21). *Extraído de www. praat. org* .

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.

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.

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.

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.

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.

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.

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.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin, 2*(6), 110-114.

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.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (version 6.0. 21). *Extraído de www. praat. org [23/07/2017]*.

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.

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.

Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. *Strategic Management Journal, 39*(8), 2362-2385.

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.

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.

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behav Brain Sci, 27*(2), 212-225.

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.