Supplementary Material:

Impairments in probabilistic prediction and Bayesian learning can explain reduced neural

semantic priming in schizophrenia

Table of Contents

1. Use of Terms and Assumptions about Neurocognitive mechanisms

The experimental paradigm we used in the present study was designed to probe non-automatic ("controlled") semantic priming — specifically, semantic priming that is driven by relatively "implicit" probabilistic predictive mechanisms. Because we understand that terms like "automatic", "controlled", "strategic", and "implicit" are used in different ways in the literature, here we clarify how we define these terms in the present study, as well as our assumptions about the neurocognitive mechanisms probed by our paradigm.

<u>"Automatic" versus "Controlled" Semantic Priming</u>

"Automatic" semantic priming has been traditionally defined as semantic priming that is driven by mechanisms that do not require marked attention to the prime word itself, and that do not engage 'top-down' processing. The most commonly discussed automatic semantic priming mechanism is the automatic "spread of activation" across lexical and semantic networks (Neely, 1977; Posner & Snyder, 1975). On this account, after encountering a prime word (e.g., "priest"), activity flows, in a bottom-up fashion, from the prime's orthographic or phonological form (p-r-i-e-s-t) to its lexical representation (e.g. PRIEST), and then to its semantic distributed features (the semantic features associated with <priest>). These representations, in turn, automatically spread activity to shared (directly associated) semantic features (e.g. the semantic features associated with <church>). This means that if a semantically associated target word, like "church", appears very quickly after the prime, its processing is facilitated because its lexical representation and/or semantic features have already been activated, and so they are easier to access than if the target word was preceded by an unrelated prime word.

To capture this type of automatic spread of activation, the time interval between the onset

of the prime and the onset of the target (the Stimulus Onset Asynchrony, SOA) must be very short, and/or the prime word must be masked so that it is perceived under or just over the threshold of awareness (Marcel, 1983). In addition, participants should carry out tasks that encourage them to process the meaning of the target, rather than tasks like lexical decision that draw attention to the relationship between the prime and target.

In schizophrenia, the magnitude of the automatic semantic priming effect is usually same as in healthy controls (e.g., Barch et al., 1996; Blum & Freides, 1995; Chapin et al., 1992; Ober et al., 1995; Vinogradov et al., 1992). If, however, the relationship between the prime and target is *indirect* — that is, linked by an unstated mediator word (e.g. "priest" — [church] — "bell") , the automatic semantic priming effect can be larger in people with schizophrenia than in healthy controls, particularly in patients with positive thought disorder (Kreher et al., 2009; Moritz et al., 2001; Moritz et al., 2003; Spitzer et al., 1993; Weisbrod et al., 1998). This has been taken as evidence for an abnormally *broad* automatic spread of activity across semantic memory in schizophrenia. Recent evidence suggests that, rather than arising from an unconstrained, faster spread of automatic activation across semantic memory, as was originally hypothesized (Manschreck, Maher, Milavetz, Ames, Weisstein, & Schneyer, 1988), this broader automatic lexico-semantic activity in schizophrenia stems from looser *mappings* between the form and meaning of words (Kuperberg, Weber, Delaney-Busch, Ustine, Stillerman, Hämäläinen & Lau, 2019).

This type of "automatic" semantic priming can be contrasted with "controlled" mechanisms of semantic priming, which do require attention to the prime, and are conceptualized as being *top-down* in nature. They are thought to be engaged when the time interval between the onset of the prime and target is longer, and when the prime is unmasked. As

discussed in the main manuscript, under these experimental conditions, the semantic priming effect in schizophrenia is typically reduced (Pomarol-Clotet et al., 2008), and this has usually been attributed to impairments in top-down processing mechanisms (Barch et al., 1996; Barch & Ceaser, 2012). This reduced semantic priming has not been consistently associated with any specific symptom of schizophrenia, but rather appears to characterize schizophrenia as a whole.

As discussed in the main manuscript, two types of top-down "controlled" semantic priming mechanisms have been distinguished — proactive (prediction) and retroactive (semantic matching) (Neely, Keefe & Ross, 1989). The aim in the present study was to isolate the former mechanism — *predictive* semantic priming.

"Implicit" versus "Strategic" predictive mechanisms

The original models of semantic priming equated "automatic" mechanisms of priming with "implicit" mechanisms, and "controlled" top-down mechanisms with explicit "strategic" processes. By "strategic", what was usually mean was that participants were overtly aware of the mechanisms in which they were engaging to carry out the task, and that these mechanisms were intentional and conscious in nature (see Posner & Snyder, 1975). For example, top-down predictive semantic priming was assumed to involve the generation of a fixed "expectancy set" (a set of candidate targets related to the prime) after observing the prime word; then, upon presentation of the target, this expectancy set was serially searched for a word whose visual form matched the bottom-up input provided by the target word; in the event a match was found, recognition and/or processing was facilitated for the target (Becker, 1980).

More recent work, however, suggests that top-down predictive processing is not necessarily conscious and strategic in nature (although, of course, it can be subject to strategic

control). For example, there is growing understanding that predictive processes engaged during language comprehension are graded and probabilistic in nature (Federmeier & Kutas, 1999; DeLong, Urbach, & Kutas, 2005, Kutas, DeLong, & Smith, 2011, see Kuperberg & Jaeger, 2016 for a review. These observations, along with evidence supporting the role of implicit statistical learning in language learning (Saffran, Newport, & Aslin, 1996) and processing/adaptation (Kleinschmidt & Jaeger, 2015), suggest that, even under non-automatic conditions, predictive processing is likely to be more implicit and probabilistic than previously assumed.

This conceptualization of probabilistic implicit prediction forms the basis for the theoretical framework we adopt in the present study, and it is exemplified by our probabilistic model of learning/adaptation. In the present paradigm, in heathy control participants, prediction under conditions of high predictive validity was possible because participants engaged in implicit statistical learning, enabling them to adapt to the statistical structure of the contextual environment, *even though* they were never told to do so. Thus, in the present paper, we emphasize the more implicit nature of predictive semantic priming in order to clarify that we do not assume that prediction was all-or-nothing or overtly strategic in nature.

Relationship between probabilistic prediction in this semantic priming paradigm and during language comprehension

The probabilistic prediction mechanisms engaged during this semantic priming paradigm may not work in precisely the same ways as those engaged during natural language comprehension, which is obviously much more complex. However, one important advantage of this paradigm is that, by varying the proportion of predictable associated versus non-predictable unrelated trials within the broader contextual environment, we were able to encourage predictive

processes while holding the *local* semantic context on each individual trial constant. That is, trials in the lower predictive validity block had the same structure as in the higher predictive validity block; the *only* factor that varied between blocks was the relatedness proportion. This means that we cannot attribute any impairments observed in the schizophrenia group to impairments in processing the context itself (indeed, as discussed in the paper, the amplitude of the N400 evoked by the prime word did not differ between the patient and control group). Thus, any reduction in the predictive priming effect in patient group cannot easily be attributed to deficits in holding information across delays in working memory, which might, in theory, contribute to impairments in using context during higher-level sentence and discourse comprehension. However, when thinking about the role of pro-active predictive mechanisms during higher-level language comprehension, it will be important to consider the role of working memory in maintaining information in the context over longer delays.

2. Stimuli and Task

The materials used in this study were the same as those described by Weber, Lau, Stillerman, & Kuperberg (2016) and Lau, Gramfort, Hämäläinen, & Kuperberg (2013), which were very similar to the stimuli described by Lau, Weber, Gramfort, Hämäläinen, & Kuperberg (2016). Note that, while analyses in these previous studies focused on a smaller subset of stimuli, which were matched for lexical variables and counterbalanced across conditions, the analysis approach taken in the present study included all non-probe items and controlled for lexical variables by including them as covariates in our analyses. However, for the purpose of visualization, a matched subset of stimuli described in the previous studies was used to generate

6

the waveforms shown in Figure 3A. Below we first describe the full set of stimuli that were used for analysis. We then summarize the subset of stimuli used to generate the waveforms in Figure 3A.

Semantically associated word pairs (example: salt – pepper) were selected from the University of South Florida Association Norms (Nelson, McEvoy CL, & Schreiber, 2004). All associated word pairs had a forward association strength (FAS) of .32 or higher (at least 32% of participants presented with the prime word responded with the target). All primes had been normed by at least 100 participants. Unrelated word pairs were created by shuffling the targets of related word pairs; the resultant unrelated prime-target pairs were then manually checked to ensure that no associated pairs had accidentally been created (FAS = 0).

The desired relatedness proportion in each block was achieved by varying the number of associated versus unrelated pairs in each block. The lower predictive validity block was composed of 360 unrelated prime-target pairs and 40 associated prime-target pairs. The higher predictive validity block was composed of 200 unrelated prime-target pairs and 200 related prime-target pairs. No word in any position was ever repeated in a given presentation list. The order of stimuli in each list was pseudorandomized within and across participants. All primes were presented in lowercase, for 500 ms; all targets were presented in uppercase, for 900 ms. There was a 100 ms blank screen between each prime and its target, such that the stimulus onset asynchrony between prime and target was 600 ms. Each trial was followed by a 200 ms fixation cross, and then a 200 ms blank screen. (See Figure 1 in the main manuscript for a diagram of the trial structure).

Participants' task was to press a button every time they saw an animal word. For the purpose of this task, we included 'animal probe' prime-target pairs that contained animal words

(e.g., *tiger*): 50% of which were in the prime position and 50% of which were in the target position. None of these animal probe trials were included in the analyses for this study, and all were unrelated. In the lower predictive validity block, 80 of the 360 unrelated prime-target pairs were animal probe trials. In the higher predictive validity block, 80 of the 200 unrelated prime-target pairs were animal probe trials.

As noted above, for the purpose of visualization, a matched subset of stimuli described in the previous studies was used to generate the waveforms shown in Figure 3A. This matched subset of stimuli was organized in a 2 (Relatedness: associated, unrelated) x 2 (Block: lower predictive validity, higher predictive validity) design, with 40 word pairs in each of the four conditions. Note that these were the same set of stimuli that were used to carry out factorial analyses in our previous studies. These stimuli were matched across conditions for log frequency of the prime and log frequency of the target. The stimuli were also matched across blocks, within relatedness conditions, for forward association strength. Mean log frequency of the primes, for this subset was 2.55; mean log frequency of the targets was 3.53 (SUBTLEX; Brysbaert & New, 2009). Mean forward association strength for the related word pairs was .65.

In the present study, we report EEG data from patients and controls. However, these data were collected in concert with MEG data, and as part of a larger multimodal study that also included an fMRI session using the same experimental design. The order of sessions (MEG/EEG vs. fMRI) was counterbalanced across subjects, and each participant saw a completely different set of items from session 1 to session 2, but the order of the sets and the recording modality in which they appeared (EEG-MEG/fMRI) was counterbalanced across participants.

3. Recruitment and Characterization of Participants

People with a DSM-IV diagnosis of either schizophrenia or schizoaffective disorder were recruited from the Freedom Trail Clinic, which is based at the Lindemann Mental Health Center (Boston, MA, USA), an outpatient clinic and the dedicated Massachusetts General Hospital center for research and training in psychotic disorders. Diagnoses were confirmed with the Structured Clinical Interview for *DSM–III-R* (Spitzer, 1992), which was administered by a research psychiatrist. Of the 18 patients whose data we analyzed in the present study, 14 met criteria for diagnosis of schizophrenia, and the remaining 4 met criteria for a diagnosis of schizoaffective disorder. Demographically matched adult controls, without histories of psychiatric disorders (confirmed using an abbreviated version of the Structured Clinical Interview; Spitzer, 1992), were recruited by advertisement.

Written consent was obtained following the guidelines of the Massachusetts General Hospital Human Subjects Research Committee. All participants were right-handed as assessed using a modified version of the Oldfield Handedness Inventory (Oldfield, 1971). All participants were native, primarily monolingual, American English speakers. Participants were excluded if they had a history of neurological injury, medical disorders impairing neurocognitive function, or if they met DSM-IV criteria for substance abuse (within three months) or substance dependence.

Patients symptoms were assessed on the day of the experiment, or on the day of a related fMRI study (within 1-2 weeks of this study). Symptoms were assessed using the Scale for the Assessment of Positive Symptoms (SAPS; Andreasen, 1984a), and the Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1984b). Assessments were completed by a single researcher who underwent extensive training in the administration of these scales and established at least 80% inter-rater agreement with scorers in other studies, based on videotaped interviews.

Global total scores were calculated for both the SAPS and SANS by summing global ratings for each symptom cluster.

All patients were taking stable doses of antipsychotic medication. Two were prescribed typical antipsychotics, and 16 were prescribed atypical antipsychotics. One participant in the schizophrenia group was unable to report exact medication dosages; thus, the mean and standard deviation for chlorpromazine equivalents reported in Table 1 of the main manuscript are based on 17 of the 18 participants in the sample. Of the 18 patient datasets included for analysis in the present study, four were prescribed benzodiazepines, two were prescribed anticholinergics, five were prescribed SSRIs, and one was prescribed an injectable antipsychotic. We refer the reader to Table 1 in the main manuscript for further details about the samples.

Subsequent exclusions of datasets from analysis

We originally collected data from 20 patients and 22 healthy adults. However, five datasets were subsequently excluded. One control dataset was excluded due to technical problems during the recording. One schizophrenia dataset was excluded due to poor performance in the behavioral task. One schizophrenia dataset and two control datasets were excluded due to excessive EEG artifact. Importantly, exclusions due to accuracy and artifact were based on *a priori* thresholds (Accuracy: hit rate <30%; Artifact: >30% of trials removed), and all analyses were carried out after these exclusions.

4. Description of Computational Rational Adaptor Model

As discussed in the main manuscript, we used a Bayesian model to probe differences

between people with schizophrenia and healthy control participants in the computational

mechanisms by which they learned/adapted to the changed statistical structure of the contextual

environment — the higher predictive validity of Block 2 following Block 1. Below, we discuss

this model in further detail. First, we clarify the explanatory framework and terminology that we

use throughout the paper in describing and interpreting this computational model. Second, we

describe the main theoretical and mathematical principles that underlie the model.

Explanatory Framework

The model that we use in the present study is described at Marr's first level of analysis

(the computational level; Marr, 1971). That is, it explicitly specifies the abstract principles of the

computational problem to be solved, and the nature of its ideal solution — the probabilistic

generative model that describes the beliefs attributed to a 'rational' agent (Anderson, 1990).

Rather than making strong claims about 'rationality', we see the main advantage of this

type of model as allowing for explicit and precise formal *descriptions* of participants' priors,

likelihoods, and hypothesis spaces, which is necessary to be able to instantiate psychological

theory (see Tauber, Navarro, Perfors & Steyvers, 2015 for discussion). Here, our model formally

specified how participants updated their beliefs in light of new evidence, trial-by-trial, as they

transitioned from environments of lower to higher predictive validity, and how they used these

beliefs to generate probabilistic predictions about the target target, based on the prime. Using this

model, we were able to show that the schizophrenia group was less likely than the control group

to use these computational principles to influence a trial-by-trial measure of semantic processing

(the N400) as they adapted to the higher predictive validity environment.

As noted in the main manuscript, no model specified at Marr's first computational level

specifies either the precise algorithm or specific neural mechanisms that are actually used to solve the computational problem. It will therefore be important for future studies to specify practical algorithmic- or process-level inference algorithms to explain why patients failed to adapt, including algorithms that take into account limitations in cognitive resources (Anderson & Schooler, 1991; Sanborn, Griffiths, & Navarro, 2010; Griffiths, Lieder, & Goodman, 2015). It will also be important for future work to link these algorithms with precise neural mechanisms (see Yu & Cohen for an example of work that bridges across Marr's computational, algorithmic and neural levels of explanation).

Terminology

For this type of probabilistic model, there is a strong tradition in the Cognitive Sciences of using terms associated with agency (e.g. "confidence", "hypothesis" and "belief updating") to describe the underlying probabilistic computations. Crucially, this language is used even though the relevant cognitive processes are theorized to be unconscious or implicit (Perfors, Tenenbaum, Griffiths & Xu, 2011). Moreover, many researchers also adopt this terminology when describing *neural* computations within this framework. For instance, in Clark's well-known "Whatever next? Predictive brains, situated agents, and the future of cognitive science" (Clark, 2013), brains can "guess", "take on tasks", "build models", "make attempts", "have confidence", et cetera….

Importantly, the use of this language does *not* imply that the brain literally "believes" or has "confidence". Rather, as discussed by the philosopher, Daniel Dennett (the Intentional stance, 1971; capital 'I' to differentiate from a more general notion of intentionality), ascribing beliefs and desires to the behavior of systems for which we do not know the precise functional design (like the brain), can allow us to better understand, explain, and predict the behavior of

these systems. See McGregor's "The Bayesian stance: Equations for 'as-if' sensorimotor agency" (2017) for a recent in-depth discussion of how a Dennettian approach of this sort can be used to explain the behavior of physical and biological systems.

Rational probabilistic model of trial-by-trial adaptation

Full details of the model we used are given in Delaney-Busch, Morgan, Lau & Kuperberg, 2019. Here we describe its main principles.

To describe a participant's beliefs about the probability of seeing an associated word pair versus an unrelated pair, the model assumes a beta-binomial distribution over trial types (associated, unrelated). The beta-binomial distribution is parametrized with a mean, μ, and a precision, ν. The mean parameter, μ, is used to estimate the probability with which a participant expects to receive an associated trial. The precision parameter (also known as the concentration parameter) describes the agent's confidence in this belief, and its prior effectively determines *how quickly* participants adapt to a new experimental environment. It can be thought of as the "sample size", or the weight given to the prior observations in 'pseudocounts'. For example, if ν = 20, then participants give the same weight to 20 trials of new data as to their prior beliefs.

At the beginning of Block 2, we set a prior of $\mu = 0.1$ — that is, we assumed that participants entered Block 2 with a belief that the parameters of the experiment would be the same as in Block 1, i.e., a 10% chance of receiving a related trial. Based on our previous study (Delaney-Busch, et al., 2019), we set a prior of $\nu = 77$. With $\mu = 0.1$ and $\nu = 77$, the beta prior at the beginning of Block 2 can alternatively be expressed in terms of the pseudocount parameterization, $\mu = \alpha/(\alpha+\beta)$ and $\nu = \alpha+\beta$ — that is, beta(7.7, 69.3), or 7.7 pseudocounts of related trials and 69.3 of unrelated trials. After each new prime-target pair in Block 2, the beta

13

distribution describing the agent's beliefs is updated using Bayes' rule.

Using a mixture model, this dynamically changing belief about trial type (associated or unrelated) is then used to *weight* the relative influence of two sources of long-term knowledge to generate probabilistic predictions about the target: the Forward Association Strength (FAS) from the prime (Nelson et al., 2004), and target frequency, estimated from the SUBTLEX corpus (Brysbaert & New, 2009). Thus, the model yields an output estimate of the probability of encountering every target word in Block 2. This estimate of raw probability is then log-transformed using the formula $-\log_2[\text{probability}]$, which converts it into the information theoretic measure, *surprisal* (Shannon & Weaver, 1949). See Supplementary Figure 1 below for a schematic of the model calculations.

A trial-by-trial output yielded by this rational adaptor model is computed for all targets in Block 2 in each participant, taking into account that participant's idiosyncratic history of associated and unrelated trials seen up to that point in the experiment. These trial-by-trial output values, in each participant, were then used as predictor variables in linear mixed effects regression analyses, which were designed to test the hypothesis that the model can explain human trial-by-trial N400s evoked by target words, measured over the course of Block 2.
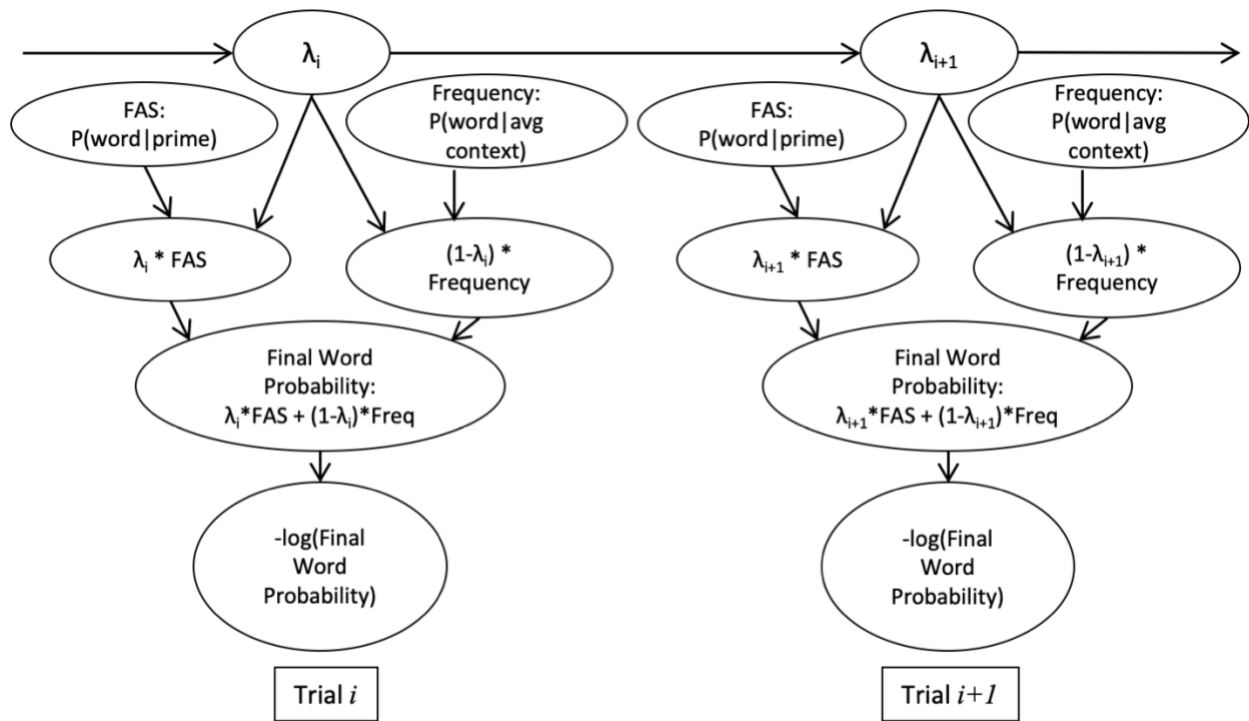
Supplementary Figure 1. Schematic of computational model of trial-by-trial Bayesian updating, see Delaney-Busch, Morgan, Lau & Kuperberg (2019) for details. λ_i: Expected probability of encountering an associated word-pair at trial i. FAS: Forward Association Strength. Freq: word frequency.

5. LMER Methods and Full Results

Below we give full results tables for the main linear mixed effects analyses (Supp. Tables 1-3) and the linear mixed effects analyses which included premorbid verbal IQ (Supp. Tables 4-6) are given below. All analyses were completed in R (R Core Team, 2016). For each analysis, we provide the model specification in R syntax, as well as the R command for running the analysis using packages lme4 version 1.1-21(Bates et al., 2015) and lmerTest version 3.1-0 (Kuznetsova, Brockhoff, & Christensen, 2017).

As noted in the manuscript, random intercepts for items and subjects were included in all models, as were random slopes for all predictors of interest that varied by item or by subject respectively. Note that, when a complex random effects structure is included in the model, often the regression analysis will not converge on an optimal solution. In the present study, we dealt with convergence issues by applying the following steps: (1) allowing for more iterations in the optimization algorithm (denoted by the *optCtrl* parameter in the lmer commands, see below), (2) changing the optimizer to the "bobyqa" optimizer provided by lme4 (Bates, et al., 2015; denoted by the *optimizer* parameter in the lmer commands, see below), and (3) excluding correlations between random intercepts and slopes from the model (denoted by the double-bar "||" notation in the model specifications, see below). All of the analyses reported in the present paper converged after applying these adjustments.

Main LMER Results

Higher Predictive Validity Block: Group*Relatedness

	Estimate (mV)	Std. Error	t-value	p-value	Sig.
(Intercept)	-0.08	0.29	-0.29	0.78	
Group	0.53	0.41	1.27	0.21	
Relatedness	0.48	0.11	4.53	0.00	***
Concreteness	0.05	0.08	0.70	0.49	
Orthographic Neighborhood Size	-0.15	0.10	-1.52	0.13	
Log Frequency	0.03	0.09	0.32	0.75	
Length	0.02	0.12	0.21	0.83	
Semantic Neighborhood Size	-0.04	0.07	-0.60	0.55	
Group*Relatedness	-0.38	0.15	-2.50	0.02	*

Supplementary Table 1: Results of linear mixed effects model in Higher Predictive Validity Block, examining modulation of the N400 evoked by target word, by Relatedness and Group. See below for model specification in R syntax.

N400_model = N400 ~ Group*Relatedness_Z + Concreteness_Z + OrthNeighb_Z + LogFreq_Z + Length_Z + SemNeighb_Z + (1 + Group*Relatedness_Z || Item) + (1 + Relatedness_Z || Group:Subject)

lmer(N400_model, HP_data, control=lmerControl(optimizer = 'bobyqa', optCtrl = list(10e4)))

Lower Predictive Validity Block: Group*Relatedness

	Estimate (mV)	Std. Error	t-value	p-value	Sig.
(Intercept)	0.09	0.29	0.30	0.77	
Group	0.14	0.42	0.34	0.73	
Relatedness	0.28	0.12	2.37	0.02	*
Concreteness	-0.12	0.08	-1.60	0.11	
Orthographic Neighborhood Size	-0.11	0.10	-1.11	0.27	
Log Frequency	0.14	0.09	1.64	0.10	
Length	0.02	0.10	0.20	0.84	
Semantic Neighborhood Size	-0.16	0.07	-2.19	0.03	*
Group*Relatedness	-0.21	0.16	-1.27	0.21	

Supplementary Table 2: Results of linear mixed effects model in Lower Predictive Validity Block, examining modulation of the N400 evoked by target words, by Relatedness and Group. See below for model specification in R syntax.

N400_model = N400 ~ Group*Relatedness_Z + Concreteness_Z + OrthNeighb_Z + LogFreq_Z + Length_Z + SemNeighb_Z + (1 + Group*Relatedness_Z || Item) + (1 + Relatedness_Z || Group:Subject)

lmer(N400_model, LP_data, control=lmerControl(optimizer = 'bobyqa', optCtrl = list(10e4)))

Effects of Bayesian Adaptor Model Output*Group (controlling for Relatedness*Group)

	Estimate (mV)	Std. Error	t-value	p-value	Sig.
(Intercept)	0.16	0.30	0.54	0.59	
Group	0.12	0.44	0.28	0.78	
Relatedness	-0.06	0.23	-0.27	0.79	
Model Output	-0.57	0.22	-2.60	0.01	**
Concreteness	0.07	0.08	0.85	0.39	
Orthographic Neighborhood Size	-0.15	0.10	-1.49	0.14	
Log Frequency	0.02	0.09	0.20	0.85	
Length	0.03	0.12	0.22	0.82	
Semantic Neighborhood Size	-0.06	0.07	-0.77	0.44	
Group*Relatedness	0.52	0.33	1.55	0.12	
Group*Model Output	0.93	0.31	3.02	0.00	**

N400_Bayes_model = N400 ~ Group*Model_Output_Z + Group*Relatedness_Z + Concreteness_Z + OrthNeighb_Z + LogFreq_Z + Length_Z + SemNeighb_Z + (1 + Model_Output_Z*Group || Item) + (1 + Model_Output_Z || Group:Subject)

lmer(N400_Bayes_model, HP_data, control=lmerControl(optimizer ="bobyqa"), optCtrl = list(10e4))

LMER Results with Premorbid Verbal IQ as a Covariate

Higher Predictive Validity Block: Group*Relatedness

	Estimate (mV)	Std. Error	t-value	p-value	Sig.
(Intercept)	0.02	0.30	0.07	0.94	
Group	0.29	0.46	0.64	0.53	
Relatedness	0.49	0.11	4.36	0.00	***
Concreteness	0.05	0.08	0.70	0.48	
Orthographic Neighborhood Size	-0.15	0.10	-1.53	0.13	
Log Frequency	0.03	0.09	0.33	0.74	
Length	0.02	0.12	0.21	0.84	
Semantic Neighborhood Size	-0.04	0.07	-0.60	0.55	
Verbal IQ	-0.27	0.23	-1.17	0.25	
Group*Relatedness	-0.40	0.17	-2.36	0.02	*
Verbal IQ*Relatedness	-0.03	0.09	-0.30	0.76	

Supplementary Table 4: Results of linear mixed effects model in Higher Predictive Validity Block, examining modulation of the N400 evoked by target word, by Relatedness and Group. Premorbid Verbal IQ (NAART; Blair & Spreen, 1989) included as a covariate. See below for model specification in R syntax.

N400_model_IQ = N400 ~ Group*Relatedness_Z + Concreteness_Z + OrthNeighb_Z + LogFreq_Z + Length_Z + SemNeighb_Z + IQ_Z*Relatedness + (1 + Group*Relatedness_Z || Item) + (1 + Relatedness_Z || Group:Subject)

lmer(N400_model_IQ, HP_data, control=lmerControl(optimizer = "bobyqa", optCtrl = list(10e4)))

Lower Predictive Validity Block: Group*Relatedness

	Estimate (mV)	Std. Error	t-value	p-value	Sig.
(Intercept)	0.06	0.31	0.20	0.84	
Group	0.20	0.47	0.41	0.68	
Relatedness	0.21	0.12	1.69	0.09	.
Concreteness	-0.12	0.08	-1.57	0.12	
Orthographic Neighborhood Size	-0.11	0.10	-1.09	0.28	
Log Frequency	0.14	0.09	1.65	0.10	.
Length	0.02	0.10	0.22	0.82	
Semantic Neighborhood Size	-0.16	0.07	-2.19	0.03	*
Verbal IQ	0.06	0.23	0.26	0.80	
Group*Relatedness	-0.06	0.18	-0.33	0.74	
Verbal IQ*Relatedness	0.17	0.09	1.88	0.06	.

Supplementary Table 5: Results of linear mixed effects model in Lower Predictive Validity Block, examining modulation of the N400 evoked by target words, by Relatedness and Group. Premorbid Verbal IQ (NAART; Blair & Spreen, 1989) included as a covariate. See below for model specification in R syntax.

*N400_model_IQ = N400 ~ Group*Relatedness_Z + Concreteness_Z + OrthNeighb_Z + LogFreq_Z + Length_Z + SemNeighb_Z + IQ_Z*Relatedness + (1 + Group*Relatedness_Z || Item) + (1 + Relatedness_Z || Group:Subject)*

lmer(N400_model_IQ, LP_data, control=lmerControl(optimizer = "bobyqa", optCtrl = list(10e4)))

Effects of Bayesian Adaptor Model Output*Group (controlling for Relatedness*Group)

	Estimate (mV)	Std. Error	t-value	p-value	Sig.
(Intercept)	0.27	0.31	0.87	0.39	
Group	-0.12	0.48	-0.26	0.80	
Relatedness	-0.06	0.23	-0.27	0.79	
Model Output	-0.60	0.22	-2.71	0.01	**
Concreteness	0.07	0.08	0.85	0.39	
Orthographic Neighborhood Size	-0.15	0.10	-1.51	0.13	
Log Frequency	0.02	0.09	0.20	0.84	
Length	0.02	0.12	0.21	0.83	
Semantic Neighborhood Size	-0.06	0.07	-0.77	0.44	
Premorbid Verbal IQ	-0.28	0.22	-1.24	0.22	
Group*Relatedness	0.52	0.33	1.55	0.12	
Group*Model Output	0.99	0.31	3.14	0.00	**
Verbal IQ*Model Output	0.07	0.08	0.91	0.36	

Supplementary Table 6: Results of linear mixed effects model in Higher Predictive Validity Block, examining modulation of the N400 evoked by target words by Model Output and Group. Premorbid Verbal IQ (NAART; Blair & Spreen, 1989) included as a covariate. See below for model specification in R syntax.

*N400_Bayes_model_IQ = N400 ~ Group*Model_Output_Z + Group*Relatedness_Z + Concreteness_Z + OrthNeighb_Z + LogFreq_Z + Length_Z + SemNeighb_Z + IQ_Z *Relatedness + (1 + Model_Output_Z*Group || Item) + (1 + Model_Output_Z || Group:Subject)*

lmer(N400_Bayes_model_IQ, HP_data, control=lmerControl(optimizer ="bobyqa", optCtrl = list(10e4))

6. Relationship between Symptoms and Semantic Priming

We had no *a priori* hypotheses about what symptoms might correlate with the N400 semantic priming effect. As noted in Supplementary Materials, section 1, although thought disorder has been in some cases associated with enhanced *automatic* semantic priming, particularly to target words that are indirectly related to their preceding prime (Spitzer et al., 1993; Weisbrod et al., 1998; Moritz et al., 2001; Moritz et al., 2003; Kreher, Goff, & Kuperberg 2009; Kuperberg et al.,

2019), the reduced priming effect observed in schizophrenia under less automatic experimental

conditions (using longer stimulus onset asynchronies between prime and target), such as in the

present study, has not consistently been linked to any specific symptom of schizophrenia.

Instead, it appears to characterize the schizophrenia group as a whole (for a review of the

behavioral literature, see Pomarol-Clotet et al., 2008; for a review of the ERP literature, see

Kuperberg, Kreher, & Ditman, 2010).

We also note that, even though our sample was suitably large to address the question we

were interested in — whether there was a significant difference between the control and the

patient group in predictive semantic priming — it was too small to find meaningful correlations

with clinical measures. Nonetheless, for completeness, we conducted exploratory analyses to

probe for potential relationships between clinical measures in the schizophrenia group and the

effect of word pair relatedness (FAS) in the Lower and Higher Predictive Validity blocks.

We examined four clinical measures: chlorpromazine equivalent, SAPS (summed global

ratings), SANS (summed global ratings), and Thought Disorder (SAPS, global rating), and we

carried out two different regression analyses for each measure: one for the Lower Predictive

Validity block and one for the Higher Predictive Validity block. The predictor of interest in each

analysis was the interaction between Relatedness (FAS) and the clinical measure being

investigated. The models included the same nuisance variables as in the main LMER analyses

reported in the manuscript (Table 2, Supplementary Tables 1-3), but we note that the results did

not change qualitatively (that is, no non-significant effects became significant) when these

nuisance variables were dropped from the model. The random effects structure in each analysis

consisted of by-subject and by-item intercepts; by-subject slopes for Relatedness; and by-item

slopes for Relatedness, the clinical variable of interest, and the interaction between the two.

Convergence issues were dealt with in the same way as in the main analyses (see Supplementary Materials, section 5).

We found no significant interactions between FAS and any of the clinical variables of interest, in either block (all ps > .142), suggesting that the N400 semantic priming effect was not modulated by any of these measures.

7. References

Anderson JR. *The Adaptive Character of Thought.* Hillsdale, NJ: Erlbaum; 1990.

Anderson JR, Schooler LJ. Reflections of the environment in memory. *Psychological Science.* 1991;2(6):396-408.

Andreasen NC. Scale for the Assessment of Negative Symptoms (SANS). In. Iowa City: The University of Iowa; 1984.

Andreasen NC. Scale for the Assessment of Positive Symptoms (SAPS). In. Iowa City, IA: The University of Iowa; 1984.

Barch DM, Ceaser A. Cognition in schizophrenia: core psychological and neural mechanisms. *Trends Cogn Sci.* 2012;16(1):27-34.

Barch DM, Cohen JD, Servan-Schreiber D, Steingard S, Steinhauer S, van Kammen D. Semantic priming in schizophrenia: An examination of spreading activation using word pronunciation and multiple SOAs. *J Abnorm Psychol.* 1996;105(4):592-601.

Bates DM, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software.* 2015;67(1):1-48.

Becker CA. Semantic context effects in visual word recognition: an analysis of semantic

strategies. *Memory and Cognition.* 1980;8(6):493-512.

Blair JR, Spreen O. Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist.* 1989;3(2):129-136.

Blum NA, Freides D. Investigating thought disorder in schizophrenia with the lexical decision task. *Schizophr Res.* 1995;16(3):217-224.

Brysbaert M, New B. 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. *Behavioral Research Methods.* 2009;41:977-990.

Chapin K, McCown J, Vann L, Kenney D, Youssef I. Activation and facilitation in the lexicon of schizophrenics. *Schizophr Res.* 1992;6:251-255.

Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci.* 2013;36(3):181-204.

Delaney-Busch N, Morgan E, Lau E, Kuperberg GR. Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition.* 2019;187(June 2019):10-20.

DeLong KA, Urbach TP, Kutas M. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat Neurosci.* 2005;8(8):1117-1121.

Dennett DC. Intentional systems. *Journal of Philosophy.* 1971;68(4):87-106.

Federmeier KD, Kutas M. A rose by any other name: Long-term memory structure and sentence processing. *J Mem Lang.* 1999;41(4):469-495.

Griffiths TL, Lieder F, Goodman ND. Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science.* 2015;7(2):217-229.

Kleinschmidt DF, Jaeger FT. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol Rev.* 2015;122(2):148-203.

Kreher DA, Goff D, Kuperberg GR. Why all the confusion? Experimental task explains discrepant semantic priming effects in schizophrenia under "automatic" conditions: Evidence from event-related potentials. *Schizophr Res.* 2009;111(1-3):174-181.

Kuperberg GR, Jaeger TF. What do we mean by prediction in language comprehension? *Lang Cogn Neurosci.* 2016;31(1):32-59.

Kuperberg GR, Kreher DA, Ditman T. What can event-related potentials tell us about language, and perhaps even thought, in schizophrenia? *International Journal of Psychophysiology: Special Issue on Language and Psychophysiology.* 2010;75:66-76.

Kuperberg GR, Weber K, Delaney-Busch N, et al. Multimodal neuroimaging evidence for looser lexico-semantic networks in schizophrenia:Evidence from masked indirect semantic priming. *Neuropsychologia.* 2019;124:337-349.

Kutas M, DeLong KA, Smith NJ. A look around at what lies ahead: Prediction and predictability in language processing. In: Bar M, ed. *Predictions in the brain: Using our past to generate a future.* New York: Oxford University Press; 2011:190-207.

Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software.* 2017;82(13).

Lau EF, Gramfort A, Hämäläinen MS, Kuperberg GR. Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *J Neurosci.* 2013;33(43):17174-17181.

Lau EF, Weber K, Gramfort A, Hamalainen MS, Kuperberg GR. Spatiotemporal signatures of lexico-semantic prediction. *Cerebral Cortex.* 2016;26(4):1377-1387.

Manschreck TC, Maher BA, Milavetz JJ, Ames D, Weisstein CC, Schneyer ML. Semantic
    priming in thought disordered schizophrenic patients. *Schizophr Res.* 1988;1:61-66.

Marcel AJ. Conscious and unconscious perception: An approach to the relations between
    phenomenal experience and perceptual processes. *Cogn Psychol.* 1983;15(2):238-300.

Marr D. Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci.*
    1971;262(841):23-81.

McGregor S. The Bayesian stance: Equations for 'as-if' sensorimotor agency. *Adaptive
    Behavior.* 2017;25(2):72-82.

Moritz S, Mersmann K, Kloss M, et al. Enhanced semantic priming in thought-disordered
    schizophrenic patients using a word pronunciation task. *Schizophr Res.* 2001;48(2-
    3):301-305.

Moritz S, Woodward TS, Kuppers D, Lausen A, Schickel M. Increased automatic spreading of
    activation in thought-disordered schizophrenic patients. *Schizophr Res.* 2003;59(2-
    3):181-186.

Neely JH. Semantic priming and retrieval from lexical memory: Roles of inhibitionless
    spreading activation and limited-capacity attention. *Journal of Experimental Psychology:
    General.* 1977;106:226-254.

Neely JH, Keefe DE, Ross K. 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.* 1989;15(6):1003-1019.

Nelson DL, McEvoy CL, Schreiber TA. The University of South Florida free association, rhyme,
    and word fragment norms. *Behavior Research Methods, Instruments, & Computers.*
    2004;36(3):402-407.

Ober BA, Vinogradov S, Shenaut GK. Semantic priming of category relations in schizophrenia. *Neuropsychology.* 1995;9(2):220-228.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia.* 1971;9(1):97-113.

Perfors A, Tenenbaum JB, Griffiths TL, Xu F. A tutorial introduction to Bayesian models of cognitive development. *Cognition.* 2011;120(3):302-321.

Pomarol-Clotet E, Oh TM, Laws KR, McKenna PJ. Semantic priming in schizophrenia: Systematic review and meta-analysis. *Br J Psychiatry.* 2008;192(2):92-97.

Posner MI, Snyder CRR. Attention and cognitive control. In: Solso RL, ed. *Information Processing and Cognition: the Loyola Symposium.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1975:55-85.

R: A language and environment for statistical computing [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science.* 1996;274(5294):1926-1928.

Sanborn AN, Griffiths TL, Navarro DJ. Rational approximations to rational models: alternative algorithms for category learning. *Psychol Rev.* 2010;117(4):1144-1167.

Shannon CE, Weaver W. *The Mathematical Theory of Communication.* Urbana: University of Illinois Press; 1949.

Spitzer M, Braun U, Hermle L, Maier S. Associative semantic network dysfunction in thought-disordered schizophrenic patients: Direct evidence from indirect semantic priming. *Biol Psychiatry.* 1993;34:864-877.

Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R

(SCID). I: History, rationale, and description. *Arch Gen Psychiatry.* 1992;49(8):624-629.

Tauber S, Navarro D, Perfors A, Steyvers M. Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychol Rev.* 2017;124(4):410-441.

Vinogradov S, Ober BA, Shenaut GK. Semantic priming of word pronunciation and lexical decision in schizophrenia. *Schizophr Res.* 1992;8:171-181.

Weber K, Lau EF, Stillerman B, Kuperberg GR. The Yin and the Yang of Prediction: an fMRI study of semantic predictive processing. *PLOS One.* 2016;11(3).

Weisbrod M, Maier S, Harig S, Himmelsbach U, Spitzer M. Lateralised semantic and indirect semantic priming effects in people with schizophrenia. *Br J Psychiatry.* 1998;172:142-146.

Yu AJ, Cohen JD. Sequential effects: Superstition or rational behavior? Paper presented at: Advances in Neural Information Processing Systems 21 (NIPS 2008)2008; Vancouver, BC.