

Using Computational Patients to Evaluate Illness Mechanisms in Schizophrenia

Ralph E. Hoffman, Uli Grasemann, Ralitz Gueorguieva, Donald Quinlan, Douglas Lane, and Risto Miikkulainen

Background: Various malfunctions involving working memory, semantics, prediction error, and dopamine neuromodulation have been hypothesized to cause disorganized speech and delusions in schizophrenia. Computational models may provide insights into why some mechanisms are unlikely, suggest alternative mechanisms, and tie together explanations of seemingly disparate symptoms and experimental findings.

Methods: Eight corresponding illness mechanisms were simulated in DISCERN, an artificial neural network model of narrative understanding and recall. For this study, DISCERN learned sets of autobiographical and impersonal crime stories with associated emotion coding. In addition, 20 healthy control subjects and 37 patients with schizophrenia or schizoaffective disorder matched for age, gender, and parental education were studied using a delayed story recall task. A goodness-of-fit analysis was performed to determine the mechanism best reproducing narrative breakdown profiles generated by healthy control subjects and patients with schizophrenia. Evidence of delusion-like narratives was sought in simulations best matching the narrative breakdown profile of patients.

Results: All mechanisms were equivalent in matching the narrative breakdown profile of healthy control subjects. However, exaggerated prediction-error signaling during consolidation of episodic memories, termed hyperlearning, was statistically superior to other mechanisms in matching the narrative breakdown profile of patients. These simulations also systematically confused autobiographical agents with impersonal crime story agents to model fixed, self-referential delusions.

Conclusions: Findings suggest that exaggerated prediction-error signaling in schizophrenia intermingles and corrupts narrative memories when incorporated into long-term storage, thereby disrupting narrative language and producing fixed delusional narratives. If further validated by clinical studies, these computational patients could provide a platform for developing and testing novel treatments.

Key Words: Artificial neural network, delusions, derailment, memory, narrative language, prediction-error, schizophrenia

Certain language behaviors are characteristic of schizophrenia. Spoken discourse often fails to express a cohesive message (1–4). Many patients express fixed delusions as spurious narrative schema repeated over time intervals ranging from weeks to years (5,6).

Mechanisms causing these behaviors remain uncertain. Language disorganization has been associated with disrupted working memory, semantic processing, attention, and linguistic context (7–21), while delusions have been associated with aberrant emotion-based reasoning and associative learning, anomalous perceptions, and jumping to conclusions (22–25). Both syndromes have been associated with disturbed executive functioning (22,26,27), theory of mind (22,28,29), and dopamine neuromodulation (30,31) and elevated hippocampal/parahippocampal activation (32–34). In this situation, connectionist models employing artificial neural networks can be used to compare the likelihood of alternative mechanisms and tie together explanations of seemingly disparate symptoms and experimental findings.

Connectionist models have been used to simulate some cognitive impairments associated with schizophrenia (35–39) but not

their characteristic narrative language. Below, we describe the generation of stories by an established connectionist model called DISCERN (40–42). Its details, many of which are not essential in understanding this study, can be found in the literature and in Supplement 1. Key aspects of DISCERN, however, can be summarized as follows: 1) DISCERN learns to recognize words and the sentences and stories incorporating them using interconnected neural network modules dedicated to these different language processing levels; 2) modules learn by updating their internal connection strengths to minimize prediction errors while processing sequential language; and 3) after a group of stories is learned, DISCERN can recall any single story when prompted with an initial segment. In this study, several different illness mechanisms were simulated in DISCERN, and the resulting story-recall distortions were compared with those of healthy human subjects and patients with schizophrenia during a delayed story-recall task.

Methods and Materials

The DISCERN Model

DISCERN is organized as a chain of neural network modules (Figure 1). These modules communicate using neural activation patterns that represent words in semantic memory: similar word activation patterns reflect similar roles in sentences. To process a story input (Figure 1A, left), word representations are presented to the sentence parser one at a time as a sequence of activation patterns. The sentence parser builds a representation of each sentence by sequencing word representations corresponding to agent, predicate, indirect object, modifier, and direct object. At the end of each sentence, the sentence representation is passed on to the story parser. The story parser transforms sequences of sentence representations into script representations. Scripts are standardized, multi-sentence schemas whose “slots” are filled by different sets of words. A story is a sequence of scripts stored in the episodic

From the Department of Psychiatry (REH, DQ) and School of Public Health (RG), Yale University School of Medicine, New Haven, Connecticut; Department of Computer Science (UG, RM), University of Texas at Austin, Austin, Texas; and Geriatrics and Extended Care Service (DL), Veterans Affairs Puget Sound Health Care System, Tacoma, Washington.

Address correspondence to Ralph E. Hoffman, M.D., Yale University School of Medicine, Yale-New Haven Psychiatric Hospital, Department of Psychiatry, 184 Liberty Street, New Haven, CT 06519; E-mail: ralph.hoffman@yale.edu.

Received Oct 27, 2010; revised Dec 23, 2010; accepted Dec 23, 2010.

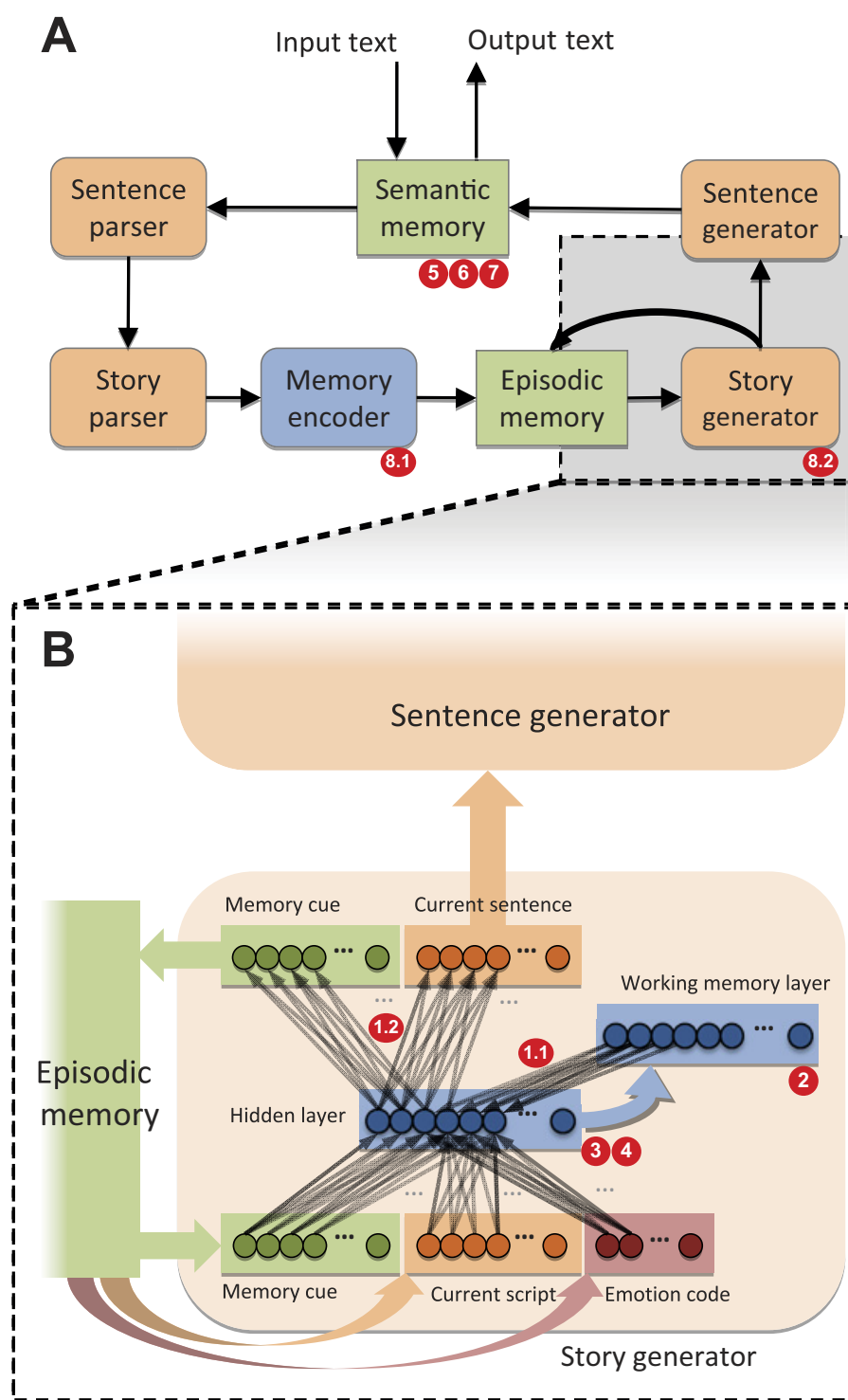


Figure 1. A schematic representation of the architecture of DISCERN. **(A)** Remembering and reproducing a story in DISCERN is achieved by a chain of modules (40–42). Tan-colored modules are simple recurrent networks (43). **(B)** The story generator simple recurrent network module shown in more detail. Hidden layer and recurrent layer interactions constitute a working memory (WM). The modules in DISCERN communicate using distributed representations of word meanings, i.e., fixed-size patterns of neuron activations, stored in a central lexicon. These representations are learned based on how the words are used in the example stories, using a modified version of backpropagation. The memory trace for each story was a compacted representation of a sequence of scripts and their slot fillers. Alternative illness mechanisms simulated in DISCERN include: 1.1) WM network disconnection; 1.2) disconnection extended to include hidden → output layer story generator connections; 2) WM noise; 3) WM gain reduction; 4) lowered WM neural bias simulating elevated arousal; 5) semantic network distortion; 6) excessive semantic network activation; 7) heightened semantic priming; 8.1) hyperlearning simulated as exaggerated prediction-error signal during memory consolidation in the memory encoder; and 8.2) hyperlearning extended to the story generator module. Details regarding the architecture, function, and training of DISCERN are provided in Supplement 1.

memory module with coherent slot fillers. To generate an output story, the process is reversed (Figure 1A, right): the story generator module accesses the episodic memory representation and translates it into a sequence of sentences. The sentence generator then produces the word sequence for each sentence.

In this study, DISCERN learned 28 stories, half autobiographical, describing a first-person character, a doctor, his relationships with boss, girlfriend, etc., and activities such as going to a wedding and receiving a traffic ticket. The other half were impersonal crime stories describing Mafia, police characters, and their activities. The lexicon was 159 words. Each story was organized as a sequence of three to seven scripts slotted with a coherent set of words/concepts and an emotion code ranging from very positive to very negative (+/+/-/-/-/-). Memory recall is thus biased by emotionality analogous to human memory retrieval (44). Such emotionality could heighten simulated derailments and delusional language, as it does in patients (22,45). Autobiographical and crime stories at times incorporated common scripts, thereby providing opportunities for narrative confusion. For instance, an autobiographical story incorporated the following script (underlining indicates slot fillers):

I was a doctor
I worked in New-York
I liked my job
I was good doctor

whereas a crime story expressed the same script with different slot fillers:

Tony was a gangster
Tony worked in Chicago
Tony hated his job
Tony was a bad gangster

Story learning in DISCERN is based on discrepancies between observed and predicted language. Such discrepancies, or prediction errors, are propagated back from the output to the input neuron layers within each module (Figure 1B); through the backpropagation learning algorithm, they produce gradual, highly targeted changes in network connection strengths (46). This learning process requires thousands of repetitions. Successful learning for this study required between 5,000 and 30,000 backpropagation learning cycles for each module. Starting from different initial random connection weights, 30 DISCERN exemplars were independently developed in this fashion.

Story recall by each DISCERN exemplar was prompted by the first script of each story as input. DISCERN parsed this script, retrieved a story memory, and produced a story output.

Eight candidate illness mechanisms were applied to each of the 30 DISCERN exemplars after story learning was completed (Figure 1), based on prior studies of speech disorganization, delusions, and schizophrenia:

1. Working memory (WM) disconnection was prompted by neuroimaging studies suggesting cortical disconnection, especially involving WM networks, in schizophrenia (47–50). Disconnection was simulated by pruning excitatory and inhibitory WM connections in the story generator if their absolute connection strength fell below a specified threshold (51). An extended version of disconnection also pruned connections between the hidden → output layer of the story generator.
2. Noise added to WM networks was prompted by reports indicating excessive cortical noise, reduced signal-to-noise ratio, and inefficiency in frontal WM cortical systems in persons with

schizophrenia (52–54). These conditions were simulated by adding Gaussian noise to story generator WM neuron outputs.

3. Working memory network gain reductions were prompted by a connectionist model of hypodopaminergic cortical neuro-modulation in patients with schizophrenia expressed as reduced neural response (36) and neuroimaging studies showing reduced activation in WM circuits during task performance in patients with disorganization symptoms (55,56). This alteration was modeled as reduced gain (i.e., slope) of the response curve of neurons in the hidden and recurrent layers of the story generator module (36).
4. Response bias shifts. Elevated arousal, which could produce overactivation at neuronal level, was simulated as lateral shifts in the response curve of WM layer story generator neurons (57). This manipulation, in theory, could also simulate a failure to deactivate the superior temporal gyrus when performing a WM task detected in early-phase patients with schizophrenia (58).
5. Semantic network distortion demonstrated by lexical categorization, priming, and fluency tasks has been statistically linked to language disorganization and schizophrenia (11–13). These abnormalities were simulated by adding noise to word representations in the semantic memory.
6. Excessive activation in semantic networks. Increased temporal and prefrontal activity during semantic associations and increased activation of the cingulate cortex during object naming has been reported in schizophrenia (14,59). These disturbances were simulated by increasing output activation of neurons in the semantic network.
7. Heightened semantic priming. Studies have suggested heightened spread of activation in semantic networks in patients with schizophrenia based on word association data, especially among patients with language disorganization (10,15–17). This disturbance was simulated by blurring semantic network outputs so that words semantically linked to a target word were co-activated.
8. Exaggerated prediction-error signaling (hyperlearning). Elevated brain response to prediction error during learning has been linked statistically to delusion formation (25). Moreover, prediction-error coding and other salience-driven aspects of learning appear to be mediated, at least in part, by dopaminergic pathways (60–62); elevated dopamine release is associated with schizophrenia (31). Backpropagation learning in DISCERN is driven by prediction-error. Exaggerated prediction-error signaling was consequently represented as amplified backpropagation learning rates, termed hyperlearning, applied for 500 backpropagation learning cycles to the memory encoder after DISCERN was trained. An extended version of hyperlearning also applied this mechanism to the story-generator module (Figure 1A).

Based on evidence of an editor function during human speech production (63), an output sentence filter was incorporated into the story generator that estimates the distortion of a sentence as the average computational distance between each component word of a sentence representation and its nearest lexical template in semantic memory. If the distortion exceeds a certain threshold, the sentence is discarded. The filter reduces word selection errors and disorganized language at the cost of reducing successful recall.

Human Story Recall

Story-recall performance data from 20 normal subjects and 37 outpatients with schizophrenia or schizoaffective disorder were compared. All subjects provided written informed consent to par-

Table 1. Illustration of Method for Assessing Story Recall for First Two Sentences of “The Gift”

Proposition List ^a	Subject Recall: “I remember a whispering man that had flowers on a bus and he saw a girl and she wanted them . . .” ^b
(i) A (man) sat in a seat on the bus 1. a man rode or is on a bus .5. there was a man in some sort of vehicle	1 (“man on a bus . . .”)
(ii) man was a wispy/old 1. old man + indication of frailty .5. old man or frail man	0
(iii) (man) was hold a bunch of flowers 1. (man) possessed, holding or carrying flowers .5. (man) possessed something	1 (“[man] had flowers”)
(iv) A young girl was/sat across the aisle from the man 1. female sitting next to, near, or across from man .5. female riding in the same vehicle as man	.5 (“he saw a girl”)
(v) The girl’s eyes came back again and again to the man’s flowers. 1. female paid special attention to the flowers .5 female noticed or wanted something	.5 (“[girl] wanted [the flowers]”)

^aPropositional breakdown on the left with criteria for full and partial scores.
^bOut of maximum score of 5, this segment assigned a score of 3. Partial credit given for (iv) because it could be inferred that the girl was riding on the bus. Only partial credit was given to (v) because there is no statement or inference that the girl paid special attention to the flowers. Insertion of term, “whispering,” received no credit and was scored as a lexical misfire (see Methods and Materials).

ticipate in the study. The DSM-IV diagnoses and symptoms were determined using the Comprehensive Assessment of Symptoms and History (CASH) (64). Patients were prospectively divided into two subgroups: those who definitely demonstrated evidence of fixed delusions expressing a plot-like, narrative scheme and those where evidence for these delusions was questionable or absent.

The experimental task consisted of three stories presented binaurally on headphones. Two of these stories, which are reproduced below, share references and themes:

“The Gift” (65)

In one seat of the bus a wispy old man sat holding a bunch of fresh flowers. Across the aisle was a young girl whose eyes came back again and again to the man’s flowers. The time came for the man to get off. He thrust the flowers into the girl’s lap. “I can see you love flowers,” he explained, “and I think my wife would like you to have them. I’ll tell her I gave them to you.” The girl accepted the flowers and watched the man get off the bus and walk through the gate of an old cemetery.

“Hitchhiker” (written for this study)

I hitched into town. A wispy old man driving a pick-up truck with his frail wife gave me a ride. I sat in the back and watched the tires kick up dust. We stopped and waited for a traffic light. I turned around and peered into the rear window. I hadn’t eaten all day and my eyes came back again and again to a bag of Fritos on the dashboard. The man got out of the truck and walked around to the back. “My wife noticed that you kept looking at the Fritos,” he explained, “and she wanted you to have them.”

The third was the “Anna Thompson” story taken from the logical memory subtest of the Wechsler Memory Scale-III (66). Immediate recall, 45-minute recall, and 7-day recall were tape-recorded and transcribed for analysis blind to group and subject identification. Seven-day recall data were used for this study.

Comparing Human and DISCERN Story Recall

Four story-recall variables could be scored comparably for both human and DISCERN story recall, while demonstrating sufficient nonzero base rates, and thus were used for assessing goodness-of-fit (GOF) of DISCERN narrative breakdowns relative to human story recall:

Recall success for humans was the total number of story propositions successfully paraphrased (scored as 1) or partially paraphrased (scored as .5) divided by the number of propositions in the target stories (= 36). Sentences in the stories were mapped into propositions, with full and partial recall for each proposition defined a priori (Table 1). For DISCERN, the total number of propositions fully reproduced was tallied and divided by the total number of propositions in the target stories (= 550). No partial scores were given because DISCERN propositions were less complex than human propositions.

Agent-slotting errors were word substitution errors involving story agents or characters. An example by a patient (recalling “The Gift”):

The girl gave the old man the flowers as a gift
reversing subject and indirect object. An example from DISCERN is

The cop arrested me for speeding
where the direct object, “me,” is substituted for the mafia character, “Vince.” Agent-slotting errors that are systematic across contexts provided a model of fixed delusions.

Lexical misfires are word or phrase substitutions not involving agents possessing sentence case roles paralleling target words or phrases that significantly change meaning. For human recall, an example of this type of recall error is (Table 1):

“wispy old man” → “whispering man”
For DISCERN, an example of a lexical misfire was:

Kate was interested in guns (substituting for books, in DISCERN stories, both terms referred to objects of interest)

Derailed clauses were text insertions comprised of an entire clause, not just a phrase or word, expressing extraneous meaning not inferable from the target story. An example generated by a control speaker when recalling “The Gift” (underlined scored as derailed clause) was:

A girl was sitting on the bus near him and he noticed her looking at his eyes.

An example while paraphrasing the same story produced by a patient is:

I remember the generosity of the flowers. I remember that he gave her flowers and she gave, she put the flowers on the tombstone. And I remember there was a truck involved, I believe it was a flatbed.

Here, three clauses were generated clearly departing from the meaning of the target story, with the last two borrowing from the “Hitchhiker” story.

For DISCERN, derailed clauses were those produced subsequent to a jump from the target story to another story in the corpus. For instance, DISCERN, after hyperlearning, initiated a story recounting a description of Joe, the boss of the first-person character, but then switched into another related story describing meeting Mary, Joe’s fiancée, at a bar:

Joe was my boss.
I hated Joe.
Joe was in his 30s.
Joe had a beard.
Joe liked baseball.
Mary was the fiancée of Joe.
I liked talking to Mary.
I gave a kiss good-bye to Mary.

Interrater reliabilities for scoring these human performance variables were in the acceptable range (R_i ’s between .75 and .98). DISCERN variables were scored algorithmically.

When comparing human and DISCERN story recall performance, agent-slotting errors, lexical misfires, and derailed clauses were recalibrated as penetrance scores, where totals for each type were divided by recall scores (before correction by total number of propositions in target texts). This strategy adjusted for: 1) much greater number of DISCERN stories compared with human stories, amplifying opportunities for error for the former; and 2) effects of editing/filtering language outputs, which reduced errors and successful recall in parallel.

Goodness-of-Fit Analysis

To assess GOF of alternative illness mechanisms, we used the following strategy: first, a four-variable story-recall profile was aver-

aged for human control subjects, and for patients. For each of the 30 DISCERN exemplars and 8 illness mechanisms, parameters were adjusted to best reproduce these two story-recall group profiles. This yielded 30 GOF measures for each mechanism and group profile. These were then used as the dependant variable in subsequent analyses of variance to identify which mechanisms best matched group profiles. Each illness mechanism was studied initially by adjusting two parameters: the mechanism parameter itself (level of noise, pruning, etc.) and the filter threshold. Goodness-of-fit used a mean square deviation metric (67). A mixed model was used with GOF as the dependent variable, the 30 DISCERN exemplars as independent subjects, and type of mechanism (8 levels) and human group (healthy control subjects vs. schizophrenia patients) as factors. Pairwise comparisons were performed to interpret differences between mean GOF of the eight illness mechanisms relative to the story-recall profile of healthy control subjects and patients. The two mechanisms best-fitting patient data (hyperlearning and WM disconnection) in two dimensions were extended by adding an additional mechanism parameter to each (Figure 1), followed by a second analysis of variance.

To model fixed delusions, DISCERN’s agent-slotting errors need to be systematic, i.e., the same confusion of an autobiographical and a crime story agent needs to recur in different output stories. Systematicity of these errors was assessed using a randomization test.

Detailed descriptions of DISCERN simulations, GOF metric, and randomization test are provided in Supplement 1.

Results

Table 2 profiles the two subject groups; Wechsler Adult Intelligence Scale (WAIS) vocabulary scores were reduced for patients compared with control subjects. Table 3 shows narrative-recall breakdown profiles for the two human subject groups. Patients were more impaired than control subjects for recall success, derailment penetrance, and agent-slotting error penetrance but not for lexical misfire penetrance. Pooling data across subjects groups and repeating analyses within groups, there was no significant correlation between any of these variables and age, parental education level, or WAIS-scaled vocabulary score.

Illness mechanisms were initially simulated in terms of a single, corresponding mechanistic parameter combined with a variable output filter for each of the 30 DISCERN exemplars. These exemplars, by parameter adjustment, were optimized to match profiles of narrative breakdown distortions of the healthy control group and the patient group, respectively. With optimized GOF as the dependent variable, a mixed model revealed a significant group \times mechanism interaction [$F(7,203) = 36.7, p < .0001$]. No mechanism had a significant advantage matching the narrative breakdown profile of control subjects [$F(7,203) = .91, p = .07$; Figure 2A]. In contrast, the eight mechanisms differed significantly in how well they matched the patients’ narrative breakdown profile [$F(7,203) =$

Table 2. Comparison of Two Groups of Subjects for Individuals Completing Day 7 Delayed Recall of Stories

	Age ^a	Gender (M/F)	Parental Education (Grades) ^{a,b}	WAIS Scaled Vocabulary Score ^a
Healthy Control Subjects ($n = 20$)	36.6 (9.0)	(11/9)	13.7 (4.0)	12.2 (3.0)
Persons with Schizophrenia ($n = 37$)	41.5 (9.6)	(16/21)	15.1 (7.6)	9.9 (4.6)
Significance Test (two-tailed)	$t(55) = 1.51, p = .14$	$\chi^2 = .72, p = .40$	$t(53) = .77, p = .44$	$t(55) = 2.04, p = .046$

F, female; M, male; WAIS, Wechsler Adult Intelligence Scale.

^aMean (standard deviation).

^bData not available for two patients.

Table 3. Comparison of Two Subject Groups on 7-Day Story Recall

	Recall Success ^a	Agent Slotting Error Penetrance ^a	Lexical Misfire Penetrance ^a	Derailment Penetrance ^a
Healthy Control Subjects (<i>n</i> = 20)	.67 (.12)	.023 (.033)	.051 (.064)	.011 (.040)
Persons with Schizophrenia (<i>n</i> = 37)	.41 (.23)	.086 (.121)	.065 (.085)	.122 (.226)
Significance Test (two-tailed, uncorrected)	<i>t</i> (55) = 4.9, <i>p</i> = .00001	<i>t</i> (44.8) = 3.0, <i>p</i> = .004 ^b	<i>t</i> (49.3) = .70, <i>p</i> = .49 ^b	<i>t</i> (50.7) = 3.8, <i>p</i> = .0004 ^{b,c}

^aMean (standard deviation).^bEqual variance not assumed.^cAfter square root transformation to normalize data.

50.5, $p < .0001$]: WM disconnection and hyperlearning were robustly superior to the other six models in terms of GOF to patient performance ($p < .0005$ on paired *t* tests) but were not different from each other (Figure 2B, Table 4).

These two best-fit mechanisms were further studied by adding another parameter to each: hyperlearning was extended to the story generator module (Figure 1A), and disconnection was extended to hidden → output layer projections in the story generator module (Figure 1B). Goodness-of-fit to the patient narrative breakdown profile improved overall [$F(1,29) = 37.3$, $p < .0001$; Figure 2C] with a significant mechanism × parameter interaction [$F(1,29) = 10.3$, $p = .003$]: three-parameter hyperlearning producing a significantly better fit to the patient narrative breakdown profile than three-parameter disconnection [$t(29) = 4.2$, $p = .0002$].

A content analysis was undertaken examining outputs of best-fit three-parameter hyperlearning simulations to the patient group. For derailments, jumps from one story to another by these simulations occurred in a highly nonrandom fashion. The same emotional valence was retained from the prederailment to the postderailment story in 90.1% of instances averaged across all 30 best-fit simulations. Moreover, autobiographical stories tended to derail to other autobiographical stories, and impersonal Mafia/police stories tended to derail to other Mafia/police stories, with only 15.1% of derailments crossing the personal/impersonal context.

Regarding delusion-like narratives, best-fit three-parameter hyperlearning exemplars produced cross-context agent-slotting errors in a highly organized fashion: the same two characters, one from an autobiographical story and the other from a crime story, were interchanged, on average, 2.4 occasions per exemplar. For example, one exemplar generated the following when recalling story 2:

Vito (substituting for Joe) was in his 30s.
Joe was a doctor.
Joe worked in New York
Joe was my boss.
I hated Joe.

Vito was the boss of the Mafia gang in the crime stories, while Joe was the boss of the first-person character in the autobiographical stories. Later, for story 10, this same simulation produced:

Vito was a famous gangster.
Vito was the boss of Tony
Tony hated Vito.
Tony feared Joe (substituting for Vito).

This confusion occurred again in recalling Story 27):

Vince went to Starbucks.
Vince sat at a table.
Vince liked Vito.
Vince feared Joe (substituting for Vito).

In DISCERN crime stories, Vince and Tony were gangsters working under Vito. A tendency to confuse Joe and Vito in human terms could lead to the emergence of a delusional belief, where a person comes to believe that his boss in the hospital is really a Mafia boss. Nonrandomness of recurrent cross-context agent-slotting confusions assessed via a randomization test was pronounced ($p < .00001$).

Table S1 in Supplement 1 shows comparable baseline characteristics of patients with and without fixed narrative delusions (FND). To test an agent-confusion model of these delusions, the number of agent-slotting errors on story recall was compared for patients with and without these delusions and healthy control subjects. As predicted, FND+ patients made significantly more agent-slotting errors compared with both other subject groups [$F(2,54) = 4.5$, $p = .015$; Duncan post hoc comparisons, $\alpha = .05$; effect size contrasting FND+ patients versus control subjects = .69; effect size contrasting FND+ patients versus FND− patients = .79]. The contrast between FND+ patients and healthy control subjects utilizing an analysis of covariance controlling for WAIS vocabulary score remained statistically significant [$F(1,44) = 7.0$, $p = .011$].

Correlations between global thought disorder ratings, medication level, number of hospitalizations, and the four story-recall variables (Table 3) computed for patients were nonsignificant.

Discussion

Whereas all eight illness mechanisms were equivalent in matching the healthy control narrative breakdown profile, hyperlearning was significantly better than the others in matching the narrative breakdown profile of patients. The differential advantage of hyperlearning in matching patient story-recall data suggests that exaggerated prediction-error signaling during memory consolidation captures pathophysiology underlying schizophrenia specifically rather than nonspecific story-recall distortions demonstrated by human subjects overall.

A majority of three-parameter best-fit hyperlearning simulations also recurrently confused specific agents in personal stories (including the self-representation) with specific agents in crime stories (and vice versa) in a highly nonrandom fashion. Noteworthy was the high frequency of agent-slotting exchanges between the hospital boss, Joe, and the Mafia boss, Vito, and parallel confusions between the “I” self-reference and underlying Mafia members, suggesting generalization of boss/underling relationships. Insofar as story scripts provide templates for assigning intentions to agents (68), a consequence of recurrent agent-slotting confusions could be assignment of intentions and roles to autobiographical characters (possibly including the self) that borrow from impersonal stories derived from culture or the media. Confusion between agent representations in autobiographical stories and those in culturally determined narratives could account for the bizarreness of fixed, self-referential delusions, e.g., a patient insisting that her father-in-law is Saddam Hussein or that she herself is the Virgin Mary. This

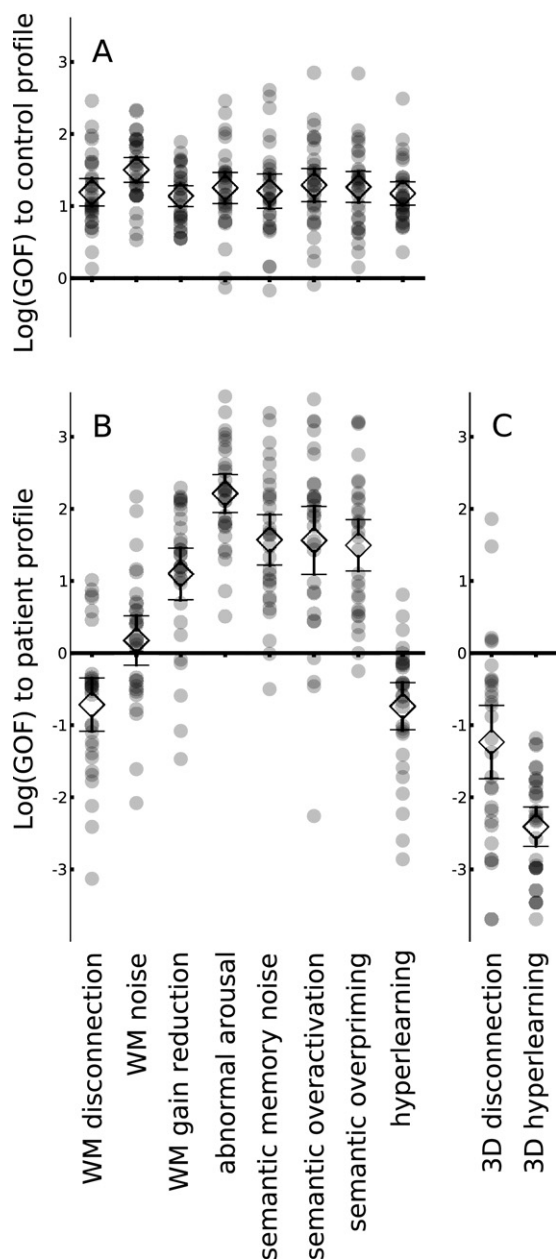


Figure 2. Scatter plots of 30 independently generated simulations showing each of the eight mechanisms mapped using a mean square deviation metric of goodness-of-fit (GOF) to the language profile of control subjects (**A**) and patients (**B**). Goodness-of-fit was log converted to normalize distributions; smaller values represent a better fit. (**C**) Adding another mechanism parameter to these disconnection and hyperlearning models improved GOF to patient performance for both, with three-dimensional hyperlearning mechanism demonstrating better GOF to patient language compared with the three-dimensional disconnection mechanism. Thus, hyperlearning emerged as the illness mechanism most likely to underlie narrative breakdown in schizophrenia. WM, working memory.

hypothesis is supported by data showing that the number of agent-slotting errors was greater in patients reporting delusions with plot-like narrative organization compared with patients without these delusions. These findings suggest that fixed delusions are story memories contaminated by misappropriated agents.

The hyperlearning model extends prediction-error abnormali-

ties in schizophrenia during associative learning (25,69,70) to learning narrative sequences stored as episodic memories. The model is consistent with a mechanistic role of excessive dopaminergic release in schizophrenia (31,71,72) insofar as dopamine release appears to enhance memory consolidation and prediction-error signaling (61,73,74). In humans, hyperlearning would likely require greater activation in hippocampal structures central to consolidating episodic memories (75–77). This formulation accounts for elevated baseline activation in hippocampal regions demonstrated in patients with schizophrenia (33,34), which appears to be reduced by antipsychotic medication (78); these medications may therefore achieve antipsychotic effects by curtailing hyperlearning, possibly by dopamine antagonistic effects (79).

Insofar as stories provide templates for understanding intentions of others (68), hyperlearning provides a model for disrupted theory of mind detected in prior studies of schizophrenia (22,28,29).

Our study has multiple limitations. First, some candidate disturbances suggested by prior studies of schizophrenia, such as dysfunctional executive control (22,26,27) and disrupted sentence-level linguistic context (20,21), were not modeled. Moreover, some semantic disturbances associated with speech disorganization (11,12) could arise from semantic network disconnection rather than noise. These disturbances will be addressed in a future iteration of DISCERN.

Second, DISCERN learns stories by backpropagating prediction-error signals (40,41,46), which are exaggerated in hyperlearning. Backpropagation learning, which requires thousands of repetitions, is unlikely to be replicated precisely in the human brain. However, long-term memory consolidation in humans also appears to be gradual and incremental, occurring over days to weeks (77). During this process, memories are replayed repeatedly (80), as they are in backpropagation (81). Therefore, hyperlearning in DISCERN may have a parallel in human narrative memory consolidation. Relatively few cycles of amplified backpropagation learning (500 epochs) were needed to match the schizophrenia narrative-breakdown profile. This finding suggests that limited bursts of hyperlearning, perhaps lasting only a few days in human terms, could produce enduring schizophrenic psychosis.

Third, thought disorder scores for patients were not significantly correlated with dependent variables developed for this study. This

Table 4. Pairwise Comparisons of Optimized GOF for Two-Dimensional Hyperlearning and WM Disconnection Relative to the Other Six Two-Dimensional Models Based on Mixed Model Analysis^{a,b,c}

	2-D Hyperlearning		2-D WM Disconnection	
	t Test	p Value	t Test	p Value
WM Noise (2)	3.9	<.0001	3.6	.0004
WM Gain Reduction (3)	7.8	<.0001	7.2	<.0001
Altered WM Bias (4)	14.7	<.0001	13.3	<.0001
Semantic Network Distortion (5)	9.9	<.0001	9.2	<.0001
Excessive Semantic Network Activation (6)	8.2	<.0001	7.7	<.0001
Semantic Blurring/Overpriming (7)	9.5	<.0001	8.8	<.0001

2-D, two-dimensional; GOF, goodness-of-fit; WM, working memory.

^aWM = working memory in story generator.

^bdf = 203, all pairwise comparisons favored 2-D hyperlearning and disconnection over other models; numbers in parentheses correspond to mechanism code illustrated in Figure 1.

^cComparison of 2-D hyperlearning versus 2-D WM disconnection in terms of optimized GOF with patient narrative breakdown profile was non-significant ($t = .09$).

could be due to the fact that these were stable outpatients demonstrating minimal evidence of impairment: mean Schedule for the Assessment of Positive Symptoms global thought disorder score fell between questionable and mild (Table S1 in Supplement 1). Including a more symptomatic, actively psychotic group may reveal correlations between story-recall distortions and thought disorder scores.

Fourth, DISCERN should be further validated by assessing its capacity to replicate patterns of story-recall failure within a healthy subject group.

These limitations notwithstanding, it is noteworthy that this is the first computational study of narrative disorganization and fixed delusional narratives characteristic of schizophrenia. Hyperlearning, hypothesized as a unitary mechanism producing both disturbances, is potentially detectable via functional neuroimaging as accelerated shifts from hippocampal to cortical representations during memory consolidation (82). If this prediction is confirmed in patients with early-phase schizophrenia, computational patients could be used to test novel somatic and psychological treatments.

This work was supported by National Institute of Mental Health Grant R01MH066228, National Institutes of Health Grant R21-DC009446, National Science Foundation Grant EIA-0303609, a Dana Foundation Grant, a National Alliance for Research on Schizophrenia and Depression Independent Investigator Award, Peterson 50th Anniversary Research Partner, and the Department of Mental Health and Addiction Services of the State of Connecticut through its support of the Abraham Ribicoff Research Center at the Connecticut Mental Health Center.

All authors report no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.

- Bleuler E (1911/1950): *Dementia Praecox or the Group of Schizophrenias*. New York: International Universities Press.
- Andreasen NC (1979): Thought, language, and communication disorders. II. Diagnostic significance. *Arch Gen Psychiatry* 36:1325–1330.
- Covington MA, He C, Brown C, Naci L, McClain JT, Fjordbak BS, *et al.* (2005): Schizophrenia and the structure of language: The linguist's view. *Schizophr Res* 77:85–98.
- Docherty NM, DeRosa M, Andreasen NC (1996): Communication disturbances in schizophrenia and mania. *Arch Gen Psychiatry* 53:358–364.
- Vinogradov S, King RJ, Huberman BA (1992): An associationist model of the paranoid process: Application of phase transitions in spreading activation networks. *Psychiatry* 55:79–94.
- Harrow M, Jobe TH (2010): How frequent is chronic multiyear delusional activity and recovery in schizophrenia: A 20-year multi-follow-up. *Schizophr Bull* 36:192–204.
- Berenbaum H, Kerns JG, Vernon LL, Gomez JJ (2008): Cognitive correlates of schizophrenia signs and symptoms: I. Verbal communication disturbances. *Psychiatry Res* 159:147–156.
- Kerns JG (2007): Verbal communication impairments and cognitive control components in people with schizophrenia. *J Abnorm Psychol* 116: 279–289.
- Kravavitz E, Dixon T, Frith C, Murray R, McGuire P (2005): Association of symptoms and executive function in schizophrenia and bipolar disorder. *Schizophr Res* 74:221–231.
- Kreher DA, Holcomb PJ, Goff D, Kuperberg GR (2008): Neural evidence for faster and further automatic spreading activation in schizophrenic thought disorder. *Schizophr Bull* 34:473–482.
- Goldberg TE, Aloia MS, Gourovitch ML, Missar D, Pickar D, Weinberger DR (1998): Cognitive substrates of thought disorder, I: The semantic system. *Am J Psychiatry* 155:1671–1676.
- Aloia MS, Gourovitch ML, Missar D, Pickar D, Weinberger DR, Goldberg TE (1998): Cognitive substrates of thought disorder, II: Specifying a candidate cognitive mechanism. *Am J Psychiatry* 155:1677–1684.
- Tallent KA, Weinberger DR, Goldberg TE (2001): Associating semantic space abnormalities with formal thought disorder in schizophrenia: Use of triadic comparisons. *J Clin Exp Neuropsychol* 23:285–296.
- Kuperberg GR, Deckersbach T, Holt DJ, Goff D, West WC (2007): Increased temporal and prefrontal activity in response to semantic associations in schizophrenia. *Arch Gen Psychiatry* 64:138–151.
- Spitzer M (1997): A cognitive neuroscience view of schizophrenic thought disorder. *Schizophr Bull* 23:29–50.
- Wentura D, Moritz S, Frings C (2008): Further evidence for “hyper-priming” in thought-disordered schizophrenic patients using repeated masked category priming. *Schizophr Res* 102:69–75.
- Moritz S, Woodward TS, Kupperts D, Lausen A, Schickel M (2003): Increased automatic spreading of activation in thought-disordered schizophrenic patients. *Schizophr Res* 59:181–186.
- Moskowitz J, Davidson M, Harvey PD (1991): Effect of concurrent distraction on communication failures in schizophrenic patients. II. Medication status correlations. *Schizophr Res* 5:153–159.
- Docherty NM, Gordinier SW (1999): Immediate memory, attention and communication disturbances in schizophrenia patients and their relatives. *Psychol Med* 29:189–197.
- Kuperberg GR, McGuire PK, David AS (1998): Reduced sensitivity to linguistic context in schizophrenic thought disorder: Evidence from on-line monitoring for words in linguistically anomalous sentences. *J Abnorm Psychol* 107:423–434.
- Kuperberg GR, Kreher DA, Goff D, McGuire PK, David AS (2006): Building up linguistic context in schizophrenia: Evidence from self-paced reading. *Neuropsychology* 20:442–452.
- Bentall RP, Rowse G, Shryane N, Kinderman P, Howard R, Blackwood N, Moore R (2009): The cognitive and affective structure of paranoid delusions: A transdiagnostic investigation of patients with schizophrenia spectrum disorders and depression. *Arch Gen Psychiatry* 66:236–247.
- Gilleen J, David AS (2005): The cognitive neuropsychiatry of delusions: From psychopathology to neuropsychology and back again. *Psychol Med* 35:5–12.
- Bell V, Halligan PW, Ellis HD (2006): Explaining delusions: A cognitive perspective. *Trends Cogn Sci* 10:219–226.
- Corlett PR, Honey GD, Aitken MR, Dickinson A, Shanks DR, Absalom AR, *et al.* (2006): Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: Linking cognition, brain activity, and psychosis. *Arch Gen Psychiatry* 63:611–621.
- Stirling J, Hellewell J, Blakey A, Deakin W (2006): Thought disorder in schizophrenia is associated with both executive dysfunction and circumscribed impairments in semantic function. *Psychol Med* 36:475–484.
- Barrera A, McKenna PJ, Berrios GE (2005): Formal thought disorder in schizophrenia: An executive or a semantic deficit? *Psychol Med* 35:121–132.
- Langdon R, Coltheart M, Ward PB, Catts SV (2002): Disturbed communication in schizophrenia: The role of poor pragmatics and poor mind-reading. *Psychol Med* 32:1273–1284.
- Frith CD, Corcoran R (1996): Exploring “theory of mind” in people with schizophrenia. *Psychol Med* 26:521–530.
- Levy DL, Smith M, Robinson D, Jody D, Lerner G, Alvir J, *et al.* (1993): Methylphenidate increases thought disorder in recent onset schizophrenics, but not in normal controls. *Biol Psychiatry* 34:507–514.
- van Os J, Kapur S (2009): Schizophrenia. *Lancet* 374:635–645.
- McGuire PK, Quested DJ, Spence SA, Murray RM, Frith CD, Liddle PF (1998): Pathophysiology of “positive” thought disorder in schizophrenia. *Br J Psychiatry* 173:231–235.
- Heckers S, Rauch SL, Goff D, Savage CR, Schachter DL, Fischman AJ, Alpert NM (1998): Impaired recruitment of the hippocampus during conscious recollection in schizophrenia. *Nat Neurosci* 1:318–323.
- Schobel SA, Lewandowski NM, Corcoran CM, Moore H, Brown T, Malaspina D, Small SA (2009): Differential targeting of the CA1 subfield of the hippocampal formation by schizophrenia and related psychotic disorders. *Arch Gen Psychiatry* 66:938–946.
- Braver TS, Barch DM, Cohen JD (1999): Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. *Biol Psychiatry* 46:312–328.
- Cohen JD, Servan-Schreiber D (1992): Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychol Rev* 99:45–77.

37. Monchi O, Taylor TJ, Dagher A (2000): A neural model of working memory processes in normal subjects, Parkinson's disease and schizophrenia for fMRI design and predictions. *Neural Netw* 13:953–973.
38. Siekmeier PJ, Hoffman RE (2002): Enhanced semantic priming in schizophrenia: A computer model based on excessive pruning of local connections in association cortex. *Br J Psychiatry* 180:345–350.
39. Carter JR, Neufeld RW (2007): Cognitive processing of facial affect: Connectionist model of deviations in schizophrenia. *J Abnorm Psychol* 116:290–305.
40. Mäkiäulainen R (1993): *Subsymbolic Natural Language Processing*. Cambridge, MA: MIT Press.
41. Mäkiäulainen R, Dyer MG (1991): Natural language processing with modular PDP networks and distributed lexicon. *Cogn Sci* 15:343–399.
42. Grasmann U, Mäkiäulainen R, Hoffman R (2007): A subsymbolic model of language pathology in schizophrenia. In: *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, 311–316.
43. Elman JL (1990): Finding structure in time. *Cogn Sci* 14:179–211.
44. Bower GH (1981): Mood and memory. *Am Psychol* 36:129–148.
45. Docherty NM, Hall MJ, Gordinier SW (1998): Affective reactivity of speech in schizophrenic patients and their nonschizophrenic relatives. *J Abnorm Psychol* 107:461–467.
46. Rumelhart D, Hinton G, Williams R (1986): Learning internal representations by error propagation. In: Rumelhart D, McClelland J, editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press, 318–362.
47. Meyer-Lindenberg A, Poline JB, Kohn PD, Holt JL, Egan MF, Weinberger DR, Berman KF (2001): Evidence for abnormal cortical functional connectivity during working memory in schizophrenia. *Am J Psychiatry* 158:1809–1817.
48. Karlsgodt KH, van Erp TG, Poldrack RA, Bearden CE, Nuechterlein KH, Cannon TD (2008): Diffusion tensor imaging of the superior longitudinal fasciculus and working memory in recent-onset schizophrenia. *Biol Psychiatry* 63:512–518.
49. Kim JJ, Kwon JS, Park HJ, Youn T, Kang DH, Kim MS, *et al.* (2003): Functional disconnection between the prefrontal and parietal cortices during working memory processing in schizophrenia: A [15(O)]H₂O PET study. *Am J Psychiatry* 160:919–923.
50. Yoon JH, Minzenberg MJ, Ursu S, Ryan Walter BS, Wendelken C, Ragland JD, Carter CS (2008): Association of dorsolateral prefrontal cortex dysfunction with disrupted coordinated brain activity in schizophrenia: Relationship with impaired cognition, behavioral disorganization, and global function. *Am J Psychiatry* 165:1006–1014.
51. Hoffman RE, McGlashan TH (1997): Synaptic elimination, neurodevelopment, and the mechanism of hallucinated "voices" in schizophrenia. *Am J Psychiatry* 154:1683–1689.
52. Potkin SG, Turner JA, Brown GG, McCarthy G, Greve DN, Glover GH, *et al.* (2009): Working memory and DLPFC inefficiency in schizophrenia: The FBIRN study. *Schizophr Bull* 35:19–31.
53. Tan HY, Callicott JH, Weinberger DR (2007): Dysfunctional and compensatory prefrontal cortical systems, genes and the pathogenesis of schizophrenia. *Cereb Cortex* 17(suppl 1):i171–i181.
54. Winterer G, Weinberger DR (2004): Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends Neurosci* 27:683–690.
55. Perlstein WM, Carter CS, Noll DC, Cohen JD (2001): Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am J Psychiatry* 158:1105–1113.
56. MacDonald AW 3rd, Carter CS, Kerns JG, Ursu S, Barch DM, Holmes AJ, *et al.* (2005): Specificity of prefrontal dysfunction and context processing deficits to schizophrenia in never-medicated patients with first-episode psychosis. *Am J Psychiatry* 162:475–484.
57. Grossberg S, Pepe J (1970): Schizophrenia: Possible dependence of associational span, bowing, and primacy vs. recency on spiking threshold. *Behav Sci* 15:359–362.
58. Crossley NA, Mechelli A, Broome F-PP, Matthiasson P, Johns LC, Johns LC, *et al.* (2009): Superior temporal lobe dysfunction and frontotemporal dysconnectivity in subjects at risk of psychosis and in first-episode psychosis. *Hum Brain Mapp* 30:4129–4137.
59. Assaf M, Rivkin PR, Kuzu CH, Calhoun VD, Kraut MA, Groth KM, *et al.* (2006): Abnormal object recall and anterior cingulate overactivation correlate with formal thought disorder in schizophrenia. *Biol Psychiatry* 59:452–459.
60. Gibbs AA, Naudts KH, Spencer EP, David AS (2007): The role of dopamine in attentional and memory biases for emotional information. *Am J Psychiatry* 164:1603–1609.
61. Iordanova MD (2009): Dopaminergic modulation of appetitive and aversive predictive learning. *Rev Neurosci* 20:383–404.
62. Niv Y, Schoenbaum G (2008): Dialogues on prediction errors. *Trends Cogn Sci* 12:265–272.
63. Fox Tree JE (2000): Coordinating spontaneous talk. In: Wheeldon L, editor. *Aspects of Language Production*. New York: Psychology Press, 375–406.
64. Andreasen NC (1987) *Comprehensive Assessment of Symptoms and History*. Iowa City: University of Iowa.
65. Cerf B (1993): The Gift. In: Carfield J, Hansen MV, editors. *Chicken Soup for the Soul*. Deerfield Beach, FL: Health Communications Inc., p. 23, reprinted with permission granted by the literary estate of Bennett Cerf. All rights reserved.
66. Wechsler D (1987): *Manual for the Wechsler Memory Scale-Revised*. San Antonio, TX: Psychological Corporation.
67. Marchiori D, Warglien M (2008): Predicting human interactive learning by regret-driven neural networks. *Science* 319:1111–1113.
68. Bower GH, Morrow DG (1990): Mental models in narrative comprehension. *Science* 247:44–48.
69. Corlett PR, Murray GK, Honey GD, Aitken MR, Shanks DR, Robbins TW, *et al.* (2007): Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. *Brain* 130:2387–2400.
70. Murray GK, Corlett PR, Clark L, Pessiglione M, Blackwell AD, Honey G, *et al.* (2008): Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol Psychiatry* 13:239, 267–276.
71. Abi-Dargham A, van de Giessen E, Slifstein M, Kegeles LS, Laruelle M (2009): Baseline and amphetamine-stimulated dopamine activity are related in drug-naïve schizophrenic subjects. *Biol Psychiatry* 65:1091–1093.
72. Howes OD, Montgomery AJ, Asselin MC, Murray RM, Valli I, Tabraham P, *et al.* (2009): Elevated striatal dopamine function linked to prodromal signs of schizophrenia. *Arch Gen Psychiatry* 66:13–20.
73. Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze HJ, Düzel E (2005): Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron* 45:459–467.
74. Schott BH, Seidenbecher CI, Fenker DB, Lauer CJ, Bunzeck N, Bernstein HG, *et al.* (2006): The dopaminergic midbrain participates in human episodic memory formation: Evidence from genetic imaging. *J Neurosci* 26:1407–1417.
75. Pieffe M, Weiss PH, Zilles K, Markowitsch HJ, Fink GR (2003): Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. *Brain* 126:650–668.
76. Chiu CY, Schmithorst VJ, Brown RD, Holland SK, Dunn S (2006): Making memories: A cross-sectional investigation of episodic memory encoding in childhood using fMRI. *Dev Neuropsychol* 29:321–340.
77. McGaugh JL (2000): Memory—a century of consolidation. *Science* 287:248–251.
78. Liddle PF, Lane CJ, Ngan ET (2000): Immediate effects of risperidone on cortico-striato-thalamic loops and the hippocampus. *Br J Psychiatry* 177:402–407.
79. Howes OD, Egerton A, Allan V, McGuire P, Stokes P, Kapur S (2009): Mechanisms underlying psychosis and antipsychotic treatment response in schizophrenia: Insights from PET and SPECT imaging. *Curr Pharm Des* 15:2550–2559.
80. Euston DR, Tatsuno M, McNaughton BL (2007): Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science* 318:1147–1150.
81. McClelland JL, McNaughton BL, O'Reilly RC (1995): Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* 102:419–457.
82. Takashima A, Nieuwenhuis IL, Jensen O, Talamini LM, Rijpkema M, Fernández G (2009): Shift from hippocampal to neocortical centered retrieval network with consolidation. *J Neurosci* 29:10087–10093.