Using Computational Patients to Evaluate Illness Mechanisms in Schizophrenia

Ralph E. Hoffman, Uli Grasemann, Ralitza 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.

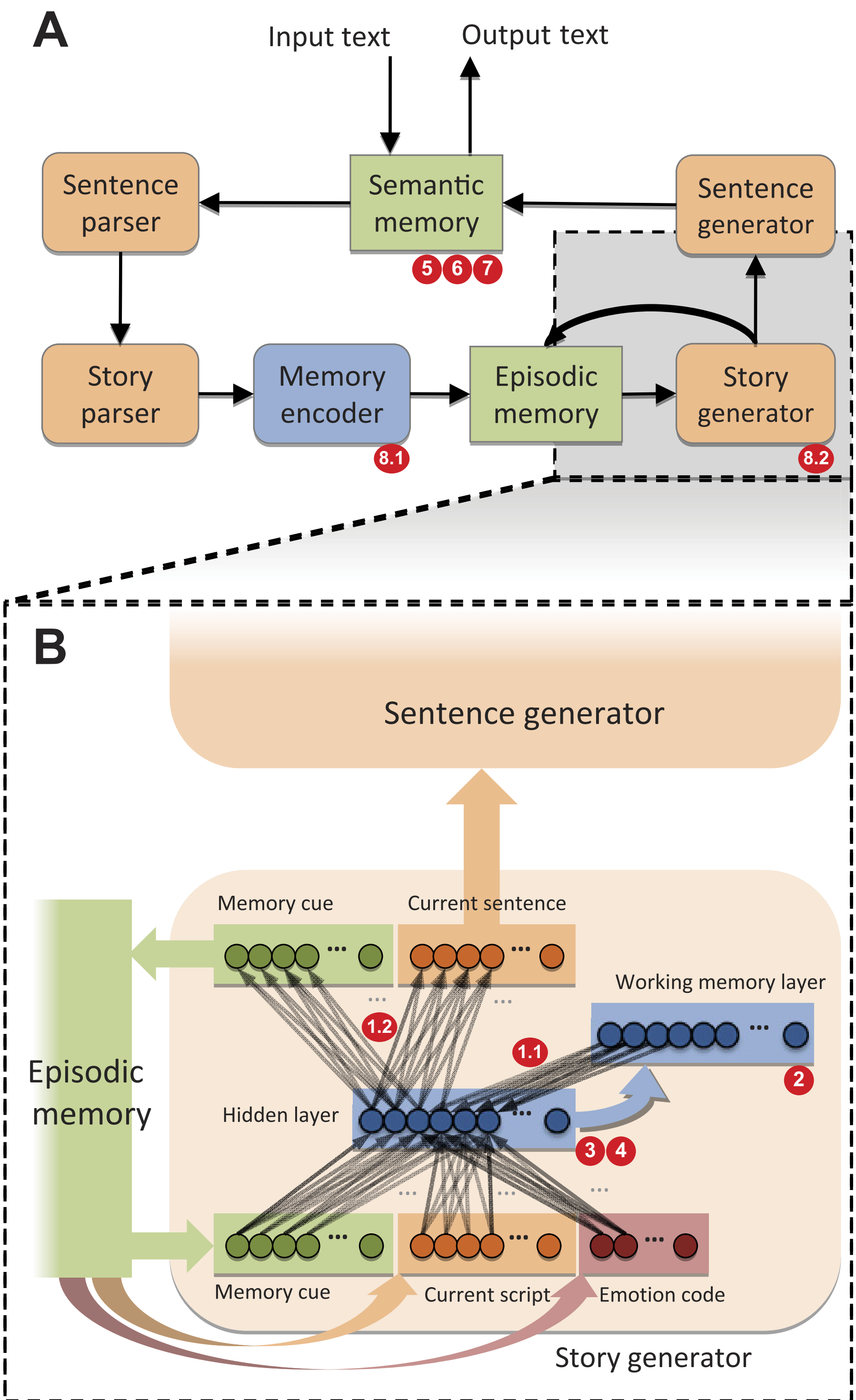

Figure 1. A schematic representation of the architecture of DISCERN. **(A)** Remembering and reproducing a story in DISCERN is achieved by a chain of modules (40–42). Tan-colored modules are simple recurrent networks (43). **(B)** The story generator simple recurrent network module shown in more detail. 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.