Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see http://www.jneurosci.org/content/jneurosci-journal-club.

An Integrative and Mechanistic Model of Impaired Belief Updating in Schizophrenia

Merage Ghane[1,2] and Tim Sparer[3]

[1]Clinical Science Program, Department of Psychology, Virginia Tech, Blacksburg, Virginia 24061, [2]Intramural Research Program, National Institute of Mental Health, Bethesda, Maryland 20892, and [3]Cognitive Psychology Program, Department of Psychology, University of Illinois Chicago, Chicago, Illinois 60607

Review of Adams et al.

Information about sensations, actions, outcomes, and their associations is often unknown, imprecise, or subject to change. Bayesian theories of learning suggest that to minimize uncertainty and maintain adaptive control, the brain optimally combines new sensory input with past experience, to infer (form beliefs about) the causes and consequences of sensations and actions (Stephan and Mathys, 2014). The degree to which beliefs are updated to support exploitation of past knowledge or sensitivity to new sensory information depends on the strength of that information and associated confidence.

Hallucinations (false perceptions) and delusions (persistent false beliefs; i.e., positive symptoms) are core symptoms of schizophrenia and are accompanied by chronic impairments in motivational/affective (negative symptoms) and cognitive domains (Fig. 1; American Psychiatric Association, 2013). Using a Bayesian framework, it is theorized that positive symptoms in schizophrenia might arise

from impaired belief updating (Stephan and Mathys, 2014). Some hypothesize that these impairments are because of overexcitation at the sensory level such that irrelevant or noisy sensory input improperly enters awareness and influences beliefs (for review, see Valton et al., 2017). Others posit that atypical belief updating is because of higher-level metacognitive impairments leading to imbalance in confidence associated with new sensory input and past experiences (Frith and Friston, 2013; Valton et al., 2017). Recently, Adams et al. (2018), proposed a Bayesian "attractor-like" model that parsimoniously accounts for both types of inferential error.

Prefrontal attractor networks are a class of biologically plausible and mechanistic models of cognitive phenomena, including working memory and decision-making (for review, see Rolls et al., 2008). Repeated experience with stimuli leads to learning via adjustment of synaptic weights. Over time, this results in the formation of relatively low-energy, recurring firing patterns (stable states) that are more readily activated by input that matches pre-existing weights. Adams et al. (2018) posit that impaired belief updating in schizophrenia results from unstable attractor networks. They hypothesized that cortical networks in schizophrenia are prone to disruption from either internal

or external "noise", and thus rather than forming stable recurrent states, the networks are more prone to jumping between states. Increased susceptibility to noise, in turn, decreases the likelihood of stabilization toward any one state (Rolls et al., 2008).

To test for differences in belief updating, Adams et al. (2018) used data from a version of the "Beads" task collected in two independent samples of patients and controls. Participants were shown two urns with a combination of red and blue beads; one had more red beads while the other had more blue beads. The urns were then hidden, and beads were sequentially drawn from one of the urns (with replacement). Participants rated the subjective probability that each presented bead was drawn from one of the urns (continuous rating from 0 to 100 in Dataset 1 and Likert rating from 1 to 7 in Dataset 2). Patients in Dataset 1 suffered from delusions, but not all were diagnosed with schizophrenia (a minority had bipolar/schizoaffective disorders). To test the specificity of findings to positive symptoms, Dataset 1 also included a non-psychotic clinical control group (with mood/anxiety disorders) from the same inpatient unit, in addition to a healthy control group. Data for all groups in Dataset 1 were collected at two time points (before and after remission). Dataset 2 in-
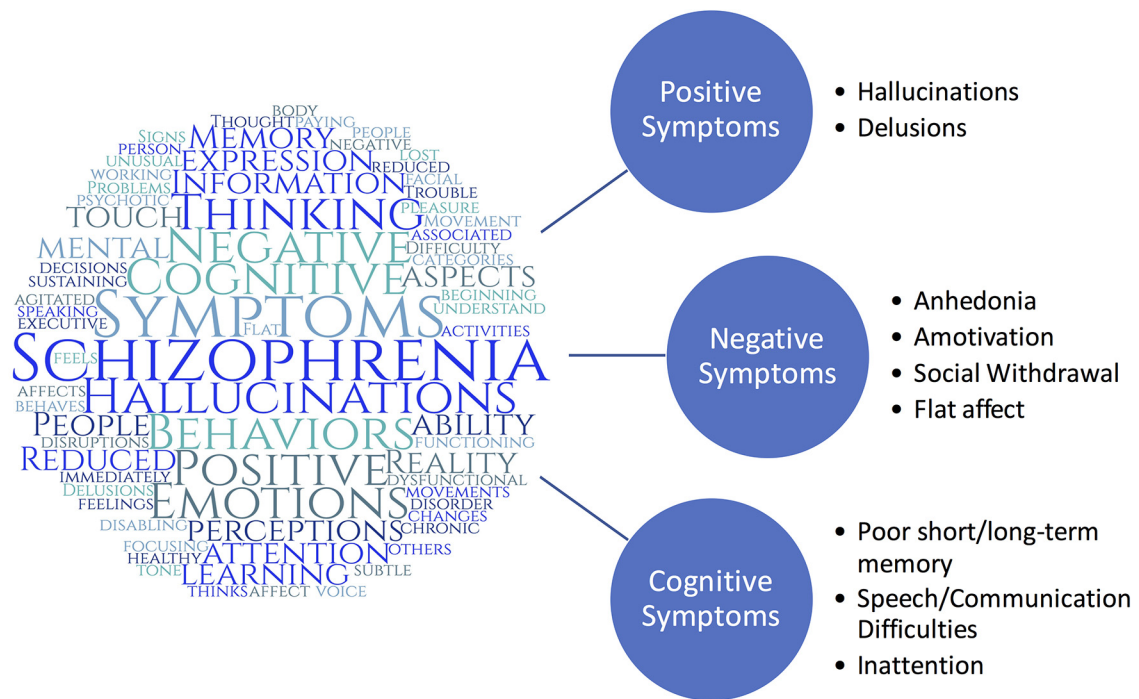
Figure 1. Left, An original diagram of the term "schizophrenia" in a word cloud created with text from the "Signs and Symptoms" section of a website published by the National Institute of Mental Health (2016)on schizophrenia. Word cloud created using web-based tool on www.WordClouds.com. Right, Examples of symptoms that fall within positive, negative, and cognitive symptom domains in schizophrenia (Rolls et al., 2008; American Psychiatric Association, 2013). Positive symptoms are thoughts, behaviors, or sensory perceptions that are atypically present in patients and outside of the sociocultural norms. Negative symptoms are thoughts, emotions, or behaviors that are absent/impaired in patients but typical of most individuals. Cognitive symptoms are undesirable changes in various cognitive domains, including memory, attention, and reasoning. For these symptoms to be of clinical relevance, they need to be persistent (at least 1 month or more) and interfere with several life domains (e.g., social relationships, work, self-care, etc.).

cluded data from a medically stable patient and healthy control group for one time point. All patients in Dataset 2 met diagnostic criteria for schizophrenia, but it was unknown whether they suffered primarily from hallucinations, delusions, or a heterogeneous mixture of these symptoms.

To test whether attractor-like dynamics accounted for individual differences in belief updating, the authors (Adams et al., 2018) defined six biologically plausible Bayesian learning models [hierarchical Gaussian filter (HGF); Mathys et al., 2011]. HGF models allow for individual-level estimation of higher-level influences (e.g., how past experiences shape sensory expectations) and lower-level inferences (e.g., how sensory input/prediction errors influence beliefs) on learning (Mathys et al., 2011). They subsequently used Bayesian model selection (Rigoux et al., 2014) to determine which model best fit data in each dataset and each group (i.e., patient and control) separately. Two of these models estimated a "belief-instability" parameter, which at high values produced simulated behavior consistent with assumptions of unstable attractor states (i.e., greater switching between states and reduced likelihood of stabilizing in any

one state). For further details on parameters included/estimated in alternative models, refer to Adams et al. (2018), their Table 3.

To estimate individual-level parameters, the authors generated a model of the task (i.e., perceptual model), followed by a model of how trial-by-trial predictions mapped onto current behavior (i.e., response model). The model that best-fit data for all groups in both datasets estimated the following parameters from the perceptual model: general learning rate (trial-to-trial variance in beliefs about urn), initial belief variance, and belief instability (strength of belief update to new sensory input when individual is uncertain versus confident in their predictions). To account for alternative sources of noise (e.g., random neuronal firing not captured by environmental variability), the authors also estimated a response stochasticity parameter for all response models (inverse variance in trial-to-trial mapping of predictions to responses).

Statistical tests in Dataset 1 revealed that both patients suffering from delusions and clinical controls showed greater belief instability and response stochasticity than healthy controls, with no difference between patients and clinical

controls. After remission, patients suffering from delusions still had greater belief instability and response stochasticity than controls, but clinical controls did not. Results were replicated in Dataset 2, with higher belief instability and response stochasticity in patients with schizophrenia than controls. In Dataset 2, initial belief variance was different between patients and controls, but this difference was no longer statistically significant after comparing patients with a subset of control participants better matched on age and sex. Belief instability and response stochasticity parameters were moderately and reliably correlated in both datasets and in simulations, suggesting that higher belief instability was associated with a weaker link between expectations and responses.

Behavioral findings reported by Adams et al. (2018) largely replicated existing literature on probabilistic decision-making and learning in schizophrenia (Peters and Garety, 2006; Corlett et al., 2009; Averbeck et al., 2011; Jardri et al., 2017). Compared with controls, patients with schizophrenia overweighted unexpected evidence, underweighted expected evidence, and were more confident when sensory evidence was limited (i.e., jumped

to conclusions). Computational modeling results provided support for unstable attractor-like dynamics as a plausible mechanism for this range of belief-updating impairments. Specifically, at high uncertainty, patients with schizophrenia were more likely to switch beliefs upon observing new sensory input (high belief instability) and were less likely to stabilize beliefs toward either urn as evidence increased (high response stochasticity). Importantly, this model parsimoniously accounted for both metacognitive and sensory interpretations of these inferential impairments. The authors also partially ruled out other possible explanations (e.g., no differences on initial belief variance or general learning rate), suggesting that unstable attractor-like dynamics accounted for behavioral differences between patients and controls.

Existing literature has largely assumed that belief updating impairments in schizophrenia are specific markers of acute positive symptoms or vulnerability to delusions (Garety and Freeman, 2013). For example, Woodward et al. (2009) showed that in patients with schizophrenia, decreases in delusion-related symptoms from precognitive to postcognitive behavioral therapy corresponded with increased evidence gathering (less jumping to conclusions). Notably, this effect was not dependent on working memory demands or the reward value of a correct response, leading the authors to conclude their specificity to delusion-related symptoms (Woodward et al., 2009; Garety and Freeman, 2013). Adams et al. (2018) did not test for within-group differences in belief instability or response stochasticity parameters before and after remission in Dataset 1, and only correlated symptoms before remission with parameter estimates from that time point. This makes it difficult to determine whether parameters of the unstable attractor-like model reflect change that is linked to specific symptom domains. In fact, results of the study (Adams et al., 2018) raise questions about whether impairments in belief-updating and unstable attractor dynamics, uniquely underlie positive symptoms (i.e., hallucinations and/or delusions). The authors (Adams et al., 2018) reported that before remission, clinical controls were just as impaired as actively psychotic patients in belief instability and response stochasticity parameters. Clinical controls were older, had higher IQ, higher affective symptoms (low mood/high anxiety measured via Manchester scale), and lower scores on the Proneness to Delusions Inventory (PDI) compared with patients

with schizophrenia before remission. These findings suggest that parameters of the unstable attractor-like model may capture pathological burden more generally, rather than symptoms of specific domains per se. Accordingly, Lincoln et al. (2010), found that correlations between delusion-related symptoms and impairments in belief-updating in patients with schizophrenia were mediated by IQ and negative symptom severity in separate analyses. Further research is necessary to determine the domain-specific impact of unstable attractor-like dynamics and associated model parameters, and how different treatments may influence these parameters.

Relatedly, response stochasticity was the only parameter significantly associated with symptom/cognitive measures, specifically correlating with PDI scores across the whole sample in Dataset 1 (before remission), and with negative symptoms and IQ in patients with schizophrenia in Dataset 2 (Adams et al., 2018). In addition to demographic and symptom heterogeneity between datasets, these inconsistent findings might suggest that select positive, negative, and cognitive symptoms share underlying computational and neurobiological mechanisms associated with response stochasticity. Severity in these symptom domains and difficulty forming biases consistent with past experience have been associated with altered prefrontal and striatal activity and connectivity (Gold et al., 2008; Simon et al., 2010; Arrondo et al., 2015). Recent work by Kurtz-David et al. (2019) showed that in healthy adults, severity of trial-by-trial inconsistency between expectations and responses was associated with greater activity in the ventromedial prefrontal and anterior cingulate cortex, but not the ventral striatum. These early findings suggest that noise in select prefrontal regions might represent trait-level risk for pathological behavior. However, the degree to which noise in prefrontal regions correlates with specific symptom domains may depend on context-dependent changes in neuromodulatory activity in other cortical/subcortical regions (e.g., because of chronic/acute stress, environmental changes, or medication; Keedy et al., 2009; Stephan and Mathys, 2014; Swardfager et al., 2016; Müller, 2018).

Adams et al. (2018) did not directly address the underlying neural mechanisms of unstable attractor-like dynamics but suggested that imbalance of cortical and mesolimbic NMDA/dopamine may be involved (Rolls et al., 2008; Adams et al.,

2018). Although disruption in NMDA and dopamine signaling has long been thought to contribute to schizophrenia, less prominent computational models of neuromodulator dynamics are gaining traction. For example, existing models of tonic and phasic acetylcholine–norepinephrine dynamics might supplement traditional hypotheses about unstable attractor-like dynamics in schizophrenia (Aston-Jones and Cohen, 2005; Yu and Dayan, 2005; Dayan and Yu, 2006). These neuromodulators are hypothesized to track expected and unexpected levels of stimulus-response-outcome uncertainty to balance metacognitive and sensory influences on beliefs and behavior. Therefore, imbalance and asynchrony in the activity of these neuromodulators might underlie belief instability and response stochasticity in schizophrenia. Recent work has linked these neuromodulators and their receptors to changes in inferential processing and belief-updating and symptom domains (Avery and Krichmar, 2017; Jepma et al., 2018). For example, blocking α_1-adrenergic receptors (α_1-ARs), which typically have excitatory effects on cell signaling, reduces positive symptoms in patients with schizophrenia (for review, see Maletic et al., 2017). Conversely, blocking α_2-ARs, which typically have inhibitory effects on cell signaling, reduces negative symptoms and improves cognitive functioning (Maletic et al., 2017). Cholinergic muscarinic and nicotinic receptors also contribute to the presentation of positive symptoms and play an important role in associated striatal dopamine signaling. These receptors also modulate prefrontal and hippocampal monoamine release during learning and plasticity (Hasselmo, 2006; Foster et al., 2012; Sarter et al., 2012). Attractor models of cortical cholinergic dynamics show that elevations of tonic acetylcholine can lead to impairments in associative learning and impede formation of stable attractor states, similar to those observed in schizophrenia (Kanamaru et al., 2013). This framework, therefore, may supplement existing cognitive and neuromodulator theories of symptoms in schizophrenia by providing a (simplified) computational and mechanistic index of uncertainty.

In summary, Adams et al. (2018) provided evidence for a parsimonious computational account of belief updating impairments in schizophrenia. Their model mirrors the assumptions of unstable attractor-like dynamics and provides insight into how uncertainty may shift a person's model of the environment. Fu-

ture work should test the clinical utility and specificity of belief instability and response stochasticity parameters in tracking symptom change or capturing phenotypic/mechanistic heterogeneity. The range of possible neurobiological explanations of unstable attractor dynamics, and the specificity of this effect on positive symptoms, are not explicitly addressed by this study or the HGF model. Future research should directly test the plausibility of proposed mechanisms and their relationships with computationally derived parameters *in vivo*.

References

Adams RA, Napier G, Roiser JP, Mathys C, Gilleen J (2018) Attractor-like dynamics in belief updating in schizophrenia. J Neurosci 38: 9471–9485.

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders: DSM-5. Arlington, VA: American Psychiatric Association.

Arrondo G, Segarra N, Metastasio A, Ziauddeen H, Spencer J, Reinders NR, Dudas RB, Robbins TW, Fletcher PC, Murray GK (2015). Reduction in ventral striatal activity when anticipating a reward in depression and schizophrenia: A replicated cross-diagnostic finding. Front Psychol 6:1280.

Aston-Jones G, Cohen JD (2005) Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. J Comp Neurol 493:99–110.

Averbeck BB, Evans S, Chouhan V, Bristow E, Shergill SS (2011) Probabilistic learning and inference in schizophrenia. Schizophr Res 127:115–122.

Avery MC, Krichmar JL (2017) Neuromodulatory Systems and Their Interactions: A Review of Models, Theories, and Experiments. Front Neural Circuits 11:108.

Corlett PR, Frith CD, Fletcher PC (2009) From drugs to deprivation: a bayesian framework for understanding models of psychosis. Psychopharmacology 206:515–530.

Dayan P, Yu AJ (2006) Phasic norepinephrine: a neural interrupt signal for unexpected events. Network 17:335–350.

Foster DJ, Jones CK, Conn PJ (2012) Emerging approaches for treatment of schizophrenia: modulation of cholinergic signaling. Discov Med 14:413–420.

Frith C, Friston KJ (2013) False perceptions and false beliefs: understanding schizophrenia. Neuroscience and the Human person: new perspectives on human activities. Vatican City: Pontifical Academy of Sciences.

Garety PA, Freeman D (2013) The past and future of delusions research: from the inexplicable to the treatable. Br J Psychiatry 203: 327–333.

Gold JM, Waltz JA, Prentice KJ, Morris SE, Heerey EA (2008) Reward processing in schizophrenia: a deficit in the representation of value. Schizophr Bull 34:835–847.

Hasselmo ME (2006) The role of acetylcholine in learning and memory. Curr Opin Neurobiol 16:710–715.

Jardri R, Duverne S, Litvinova AS, Denève S (2017) Experimental evidence for circular inference in schizophrenia. Nat Commun 8: 14218.

Jepma M, Brown SB, Murphy PR, Koelewijn SC, de Vries B, van den Maagdenberg AM, Nieuwenhuis S (2018) Noradrenergic and cholinergic modulation of belief updating. J Cogn Neurosci 31:1–18.

Kanamaru T, Fujii H, Aihara K (2013) Deformation of attractor landscape via cholinergic presynaptic modulations: a computational study using a phase neuron model. PLoS One 8:e53854.

Keedy SK, Rosen C, Khine T, Rajarethinam R, Janicak PG, Sweeney JA (2009) An fMRI study of visual attention and sensorimotor function before and after antipsychotic treatment in first-episode schizophrenia. Psychiatry Res 172:16–23.

Kurtz-David V, Persitz D, Webb R, Levy DJ (2019) The neural computation of inconsistent choice behavior. Nat Commun 10:1583.

Lincoln TM, Ziegler M, Mehl S, Rief W (2010) The jumping to conclusions bias in delusions: specificity and changeability. J Abnorm Psychol 119:40–49.

Maletic V, Eramo A, Gwin K, Offord SJ, Duffy RA (2017) The role of norepinephrine and its α-adrenergic receptors in the pathophysiology and treatment of major depressive disorder and schizophrenia: a systematic review. Front Psychiatry 8:42.

Mathys C, Daunizeau J, Friston KJ, Stephan KE. (2011) A Bayesian foundation for individual learning under uncertainty. Front Hum Neurosci 5:39.

Müller N (2018) Inflammation in schizophrenia: pathogenetic aspects and therapeutic considerations. Schizophr Bull 44:973–982.

National Institute of Mental Health (2016) Schizophrenia. Available at https://www.nimh.nih.gov/health/topics/schizophrenia/index.shtml.

Peters E, Garety P (2006) Cognitive functioning in delusions: a longitudinal analysis. Behav Res Ther 44:481–514.

Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014) Bayesian model selection for group studies: revisited. Neuroimage 84:971–985.

Rolls ET, Loh M, Deco G, Winterer G (2008) Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. Nat Rev Neurosci 9:696–709.

Sarter M, Lustig C, Taylor SF (2012) Cholinergic contributions to the cognitive symptoms of schizophrenia and the viability of cholinergic treatments. Neuropharmacology 62:1544–1553.

Simon JJ, Biller A, Walther S, Roesch-Ely D, Stippich C, Weisbrod M, Kaiser S (2010) Neural correlates of reward processing in schizophrenia: relationship to apathy and depression. Schizophr Res 118:154–161.

Stephan KE, Mathys C (2014) Computational approaches to psychiatry. Curr Opin Neurobiol 25:85–92.

Swardfager W, Rosenblat JD, Benlamri M, McIntyre RS (2016) Mapping inflammation onto mood: inflammatory mediators of anhedonia. Neurosci Biobehav Rev 64:148–166.

Valton V, Romaniuk L, Douglas Steele J, Lawrie S, Seriès P (2017) Comprehensive review: computational modelling of schizophrenia. Neurosci Biobehav Rev 83:631–646.

Woodward TS, Munz M, LeClerc C, Lecomte T (2009) Change in delusions is associated with change in "jumping to conclusions". Psychiatry Res 170:124–127.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681–692.