

The views expressed in this editorial are those of the author(s) and do not necessarily reflect the position of the Canadian Medical Association or its subsidiaries, the journal's editorial board or the Canadian College of Neuropsychopharmacology.

# The Bayesian brain and cooperative communication in schizophrenia

Lena Palaniyappan, MBBS, PhD; Ganesan Venkatasubramanian, MD, PhD

*The act of will and the action of the body are not two different states...; they do not stand in the relation of cause and effect but are one and the same thing.... The action of the body is nothing but the act of will .... translated into perception. – Arthur Schopenhauer, 1819<sup>1</sup>*

Our social world is an ever-changing milieu in which boundless verbal and nonverbal signals are generated by fellow humans. To ensure our survival, we must perceive certain regularities from the complexity that surrounds us. A failure to meet this daily challenge may prove costly for some; social encounters trigger several psychiatric symptoms, while social withdrawal reduces their intensity, at least temporarily.<sup>2</sup> For example, disorganization (or formal thought disorder), one of the central features of schizophrenia, presents primarily as a disruption in cooperative communication that occurs in a social context. Though many technical advances now allow us to study “socially” interacting agents in the laboratory (for example, see Kingsbury and colleagues<sup>3</sup>), psychiatric symptoms are rarely studied in the context of neural mechanisms of social encounters. To tackle this challenge, we need empirical tools to study the dynamic neural framework of social interaction, starting from a 2-person perspective. In this editorial, we first present such a tool: an emerging “active inference” perspective of cooperative communication between 2 individuals. We then introduce the 2-brain problem of formal thought disorder in schizophrenia as an exemplar case of its utility and map resulting theoretical expectations to known signs of this construct. Lastly, we highlight several experimental opportunities that arise from casting of formal thought disorder in the active inference framework.

## Brain as a Bayesian organ

One of the important pursuits of clinical neuroscience is to explain how computational failures of the brain result in symptoms and signs experienced by patients. To get here, we need a framework on how the brain is employed to compute solutions for the challenges (e.g., recognizing an object or talking to a stranger) it purports to solve. One biologically plausible unifying hypothesis casts the brain as a Bayesian

### Box 1. Definition

The term “Bayesian” refers to the process of assigning probabilities based on prior knowledge when interpreting data.

statistical organ whose function is to generate and refine models of the hidden properties of the world (i.e., causes and consequences) in order to reduce uncertainty<sup>4</sup> (Box 1). Accordingly, the brain’s models of the world are predictions (or “priors”) held by neuronal units, while perceptions are inferences that arise from the probability estimations of sensory data against models. When the senses convey data that are not explained by the held models (“prediction errors”), this leads to either further data gathering (e.g., attending to the object or acting on the environment) or a model revision to reduce further prediction errors. In this framework, actions fulfill behavioural predictions; they simply arise from the drive to reduce the error between the predicted future (prior “expectations”) and current motor states.<sup>5</sup> This idea, termed “active inference,” provides an algorithm for perceptions and actions,<sup>6</sup> which, combined with a plausible implementational sketch of neural circuits (dynamic causal models; not discussed here, but see Petzschner and colleagues<sup>7</sup> and Limongi and colleagues<sup>8</sup> for examples), opens the door for parsimonious mechanistic explanations for psychiatric symptoms.

## Precision-driven active inference

To function as a statistical engine, the brain needs to overcome 2 challenges: model selection and significance detection. First, as the sensory world is infinite, it is not plausible to instantly scan the entire model space to select a model to fit the incoming data. This is solved by a hierarchical framework of prediction propagation wherein specific model choice is dictated by the immediately higher level of predictive representations.<sup>9</sup> For example, when we perceive what is being said by the person speaking to us in a noisy bar (“Would you like one more of the same?”), our models are progressively constrained by the situation (the bar), the speaker (the bartender who knows

**Correspondence to:** L. Palaniyappan, Room 1232D, Robarts Research Institute, 100 Perth Drive, London, Ont., N6A 5K8; lpalaniy@uwo.ca

**Cite as:** *J Psychiatry Neurosci* 2022 February 8;47(1). doi: 10.1503/jpn.210231

you by name), the context (having a recently emptied glass), and the event (running a tab). Second, many a time, there is likely to be a mismatch between our expectations (prior models) and sensations (sensory data); this mismatch, termed “prediction error,” can be a “chance occurrence” due to inaccuracies in the data we gathered. The brain, as an optimal Bayesian machine, must allow only the “reliable” prediction errors to propagate further, and trigger model revision or further data acquisition (as these are expensive).

If the sensory data are noisy and highly variable, then any information in that data that we did not predict with our prior models cannot be trusted. The prediction error triggered from processing these data will be considered to be unreliable or less precise (i.e., precision is the inverse of the variance). We can say that an unreliable prediction error is given less importance (or weighting) over the prior predictions, and the prior model is retained. Similarly, our prior models themselves also have variable degrees of precision; we cannot always rely on our models. Ultimately, for every encounter with some form of incoming data, our inferential bias toward our models (priors) versus data should be dictated by the reliability (or precision) of these 2 sources. In effect, this reliability is the learned probability of success of a model from previous experiences of deploying the model. To put it simply, if the brain is a flawless statistical organ, then any “significant effect” that it detects from the gathered data must be based on both the magnitude of the signal (deviation of the data from prior model) and the reliability (inverse variance) of both the data and the model.

Several lines of evidence indicate that precision is encoded by the brain (arguably, at the synaptic level) and is used to assign weight to sensory data (in the form of prediction errors as they ascend) as well as priors (in the form of predictions that descend).<sup>10,11</sup> At times when our predictions are imprecise, for example, if a new bartender with an accent is speaking to us in a noisy environment, we either ask him to repeat what he said, move closer, tilt our head and pay more attention to his voice, or stubbornly stick to our expectations. In other words, in a volatile world where model revisions are often required, we either commit neural resources to improve the precision of sensory data (attentional selection<sup>12</sup>) or inflate the emphasis toward our predictions (confidence<sup>13</sup>). A similar process also operates for motor behaviour; when the consequences of a planned act are highly variable, we can still bias toward selecting an action by increasing precision afforded to a chosen policy or plan (i.e., increased commitment or motivation<sup>14</sup>). Note that this line of conceptualization brings us close to Schopenhauer’s assertion: action is nothing, but *will* translates to *perception*. This statement indeed foresees the perception–action loop articulated by active inference whereby action can be seen as the realization of motivated motor prediction.

The precision-driven active inference framework has been invoked in multiple transdiagnostic studies, including some published in *JPN*,<sup>15,16</sup> to explain abnormalities in perception,<sup>17</sup> interoception<sup>18</sup> and emotional expression,<sup>19</sup> to name a few. Here, we theoretically expand its application to disrupted communication in psychosis as a clinical phenomenon of interest.

## Active inference and cooperative communication

The basic premise of active inference is that our interactions with the world are attempts to reduce the uncertainty about the models we hold of it. In our social world, the most critical uncertainty is the mental state of those we interact with. Speech is a crucial act by which we interact with our social world. The process of engaging in a discourse with another person involves both perception and action; this can be understood as an active inference process jointly deployed by the speakers to decrease their respective uncertainties about each other’s mental state.<sup>20</sup> We achieve this by asking and answering questions directly to infer each other’s “model of the world.”<sup>21</sup>

For cooperative communication to occur, the 2 brains’ causal models and expectations should have a certain degree of alignment<sup>22,23</sup> (see also Kingsbury and colleagues<sup>24</sup>). Language, by its very nature, provides these shared tools at several levels of the hierarchical priors (phonetics, semantics, syntax).<sup>25</sup> Other features, such as culture, shared identity and common context, also provide higher-level alignment for shared message passing.<sup>26</sup> This sort of prior alignment is said to be species-specific and adaptive, making cooperative communication less demanding and more likely to occur in our social world.<sup>20</sup>

Brown and Kuperberg argue that predictive processes operate hierarchically across 3 levels in the generative model of language — semantic selection, syntactic construction and phonological articulation.<sup>9</sup> Thus, higher-order priors (e.g., subject of interest to be discussed) hierarchically constrain the choice of downstream action sequences (e.g., ideas, sentences, words) to achieve the desired outcome of a social discourse. The hierarchical generative process is subject to internal (comparing the lower-level selections against the discourse plan and the message, to make changes as we speak) and external monitoring (comparing intended and achieved perceptual consequences of the output) at each level. Once a discourse is initiated, the further choice of messages and the structure and the length of responses depend on this self-monitoring feedback from articulation as well as the feedback from the listener (other monitoring). Both the verbal and nonverbal outputs from the self and the other person are crucial for this feedback process.

But what prompts a change in our planned discourse? Our individual beliefs (probability estimates) about the most socially valuable message choice tune the commitment placed on our discourse plan, and this in turn affects the precision afforded to the syntactic, lexical and phonetic choices down the hierarchy that we employ to convey the message (propagation of precision). Sufficiently precise priors at each level are required for the discourse not to drift on the basis of incoming cues; this also considerably increases the speed of action selection and implementation (i.e., winning combination chosen at an optimal psychomotor speed; see Parr and colleagues<sup>27</sup> for limb movements). Through precision modulation, we afford sufficient commitment to the ongoing discourse plan, thus attenuating the relative impact of the sensory feedback. When the other person generates sufficiently precise output (for the speaker, this is feedback) that overcomes this commitment, we revise the ongoing speech production (e.g., changing the topic when we detect a reliable sign of displeasure).

In summary, cooperative human communication fits well with the active inference algorithm. Two speakers engage in a discourse to infer each other's mental state, thus reducing the uncertainty about each other. Based on their probabilistic estimates (beliefs), they choose messages, discourse structure, and lexical choices with maximum likelihood to serve the goal of "reverse inferring" the other person's mental state from the expected responses. The top-down conditional dependence of the subprocesses means that if the goal to infer the other person's mental state is imprecise, this reduces the precision of message selection and the implemented discourse plan, as we discuss below.

## Thought and language disorder in schizophrenia

Traditionally, linguistic abnormalities in schizophrenia have been considered as a part of a notable symptom cluster often termed "formal thought disorder" (FTD) or "disorganization," while later the emphasis shifted to "speech disorder."<sup>28</sup> These debates arise from a "Cartesian" approach that sees language as merely an expressive system that encodes an independently constituted thought process. Recent developments in generative linguistics have emphasized an interactive interface between thought and language; in this Un-Cartesian approach, language and thought are seen as indissociable;<sup>29</sup> this approach eschews the thought versus language versus speech disorder problem in understanding schizophrenia.<sup>30</sup> When extending to the thought versus speech disorder issue in schizophrenia, the Un-Cartesian approach is also compatible with Schopenhauer (will v. act) and the active inference framework (prediction v. behaviour).

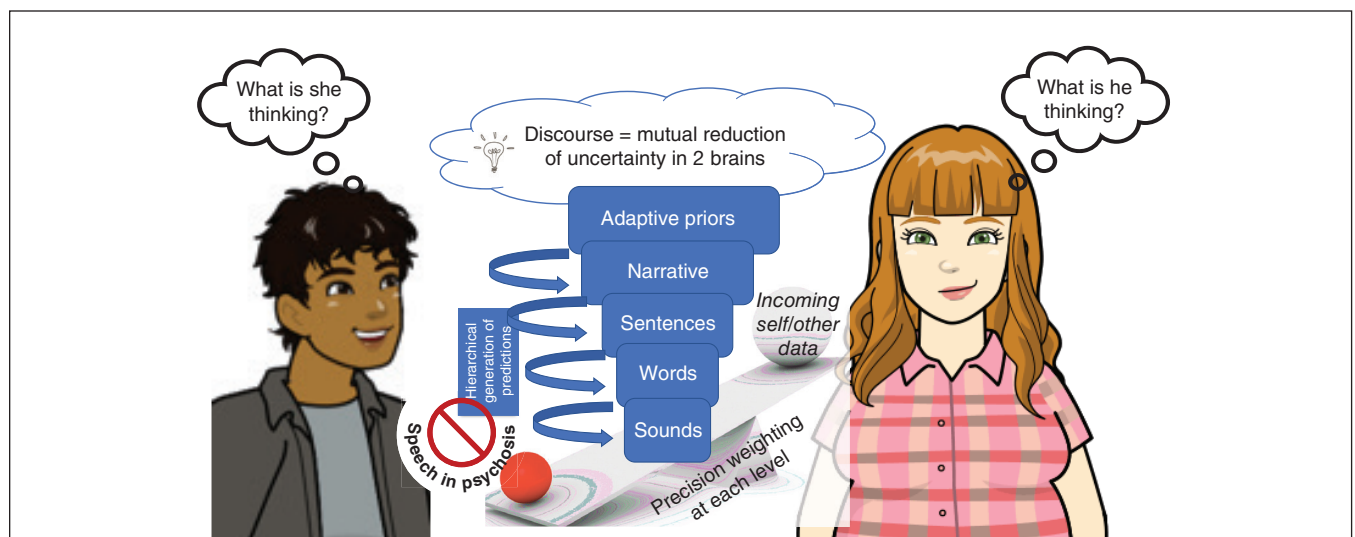
For some, FTD is essentially a 2-brain problem, and schizophrenia is primarily a disorder of communication.<sup>31,32</sup> Disrupted speech arises from the speaker using ideas, words and phrases in such a way that their meanings are not socially shared.<sup>32</sup> Several psychotic symptoms can be traced back to how the patient employs language to communicate (delusions,

hallucinations).<sup>33</sup> Nevertheless, unlike discrete language disorders such as aphasia, the hierarchical representations that form linguistic structure are affected in a reversed gradient.<sup>34</sup> In schizophrenia, the pragmatics are affected more than semantic and syntactic levels, while the phonemic choices are mostly unaffected.<sup>34,35</sup> Cohesive ties in a discourse are not preserved (positive FTD: distractibility, tangentiality, derailment, peculiar word choices) and verbal productivity diverges from contextual requirements (negative FTD: reduced speech, weakened goal, perseveration). These communication deficits play a critical role in long-term functional deficits.<sup>36,37</sup>

In an attempt to explain language comprehension and production abnormalities in schizophrenia, Brown and Kuperberg hypothesized "a breakdown of generative circuits linking high-level message representations and lexical representations" and a failure of the time-sensitive predictive processing (whereby priors are accessible only after the bottom-up inputs are fully accessed) as the basis of positive FTD in schizophrenia.<sup>9</sup> We extend these ideas to accommodate the social context of normal discourse (i.e., the 2-brain problem). In placing emphasis on 2 brains rather than 1 brain, we underscore that the neural processes pertaining to both individuals partaking in a conversation influence the "organization" or coherence. Thus, the processes that lead to verbal disorganization or FTD are best studied with both "brains" considered together, rather than seeking individual brain-symptom correlations as we practise at present.

## An active inference account of discourse in schizophrenia

Several lines of evidence argue for a deficiency in precision modulation in schizophrenia.<sup>38,39</sup> Within the hierarchical generative models of cooperative communication, we propose that a gradient of a precision modulation deficit affecting the highest levels of priors exists in patients with schizophrenia (Figure 1).



**Fig. 1.** A gradient of a precision modulation deficit affecting the highest levels of priors in schizophrenia is hypothesized to result in speech disturbances in social settings. When demands arise, patients cannot afford increased precision to their higher-order "adaptive" priors. When engaging a second person, this imprecision presents as a failure of commitment to a discourse plan (or narrative) with low confidence on the message choice that increases the likelihood of frequent shifts in conversational goal, messages and speech structure.

When demands arise, patients cannot afford increased precision to their higher-order priors. When engaging a second person, this imprecision presents as a failure of commitment to a discourse plan<sup>40</sup> with low confidence on the message choice (ambivalence). This state of low precision of higher-order priors makes all lower-level models equally likely for selection;<sup>41</sup> this increases the likelihood of frequent shifts in conversational goal, messages and speech structure (loosened associations: derailment, incoherence).

The presence of imprecise priors at various higher levels of active inference considerably reduces the speed of message selection (reduced spontaneity) and implementation (reduced rate of speech). The relatively low precision of priors compared with sensory data enables external cues to repeatedly redirect the discourse (distractibility). The between-speaker asymmetry in the precision of shared priors, and the unpredictable shifts in the discourse increases the interlocutor's uncertainty about the speaker's mental state, who loses the conversational rapport (and is left with a "Praecox feeling"<sup>42</sup>). The external cues of this uncertainty are available to the speaker, but owing to the difficulty in precision affordance to the higher-order priors, the speaker presses on with the same imprecise stream of communication (circumstantiality, empty speech with little information). Nevertheless, as the precision afforded to lower-level priors is relatively preserved, this leads to word-level errors (repetitions, perseveration, persisting with peculiar word choices) and vague references. Further, the imbalance in the dyad's relative precisions (of higher-order priors) also precludes the speakers from taking appropriate turns,<sup>23</sup> leading to conversational failures. This precision-weighting imbalance also means that, after a few exchanges, the information flow becomes lopsided. Such asymmetric exchanges arising from non-aligned priors lead to the conclusion that the patient did not infer the mental state of the other speaker.<sup>43,44</sup> In due course, the expectation that a social discourse can assist in inferring the other person's mental state diminishes greatly, as it turns out to be a costly exercise with low returns. This is especially likely if the other person also has imprecise conversational priors.<sup>45</sup> One's estimated probability of causally influencing the immediate social world reduces over time, driving down any further active inference via speech/action (diminished expressivity, negative symptoms<sup>46</sup> over the longer time course).

How does this imprecision of priors come about? First, precise priors emerge when large amounts of data have been sampled, with reinforcement of choices (termed "confidence accumulation" in decisional neuroscience<sup>47</sup>); for language, this is a developmental challenge,<sup>48</sup> which when not met, increases the chance of priors to operate with lower precision. Second, Adams and colleagues<sup>49</sup> argue a primary role for the glutamatergic system in maintaining synaptic gain and modulating precision; several empirical observations support a role for glutamatergic hypofunction in schizophrenia.<sup>50</sup> Third, it is possible that patients perceive the social environment to be more volatile,<sup>51</sup> thus estimating the consequences of their action plans to be less reliable than others. This can also arise from developmental adversities (trauma, bullying<sup>52</sup>) or experiences of significant shifts in social contexts (as in immigration). While the associations we make here are currently speculative,

experimental determination of precision weighting of different levels of priors during social conversation is entirely plausible, in line with similar studies in perceptual processing.<sup>53</sup>

Recent observations indicate that noradrenergic neuro-modulation may affect the precision of our priors;<sup>54</sup> thus, stressful states where adrenergic arousal is high may reduce reliance on (i.e., precision weighting of) priors. In contrast, dopamine may play a crucial role in precision weighting of prediction errors, rather than coding predictions per se,<sup>55</sup> while nondopaminergic modulators such as acetylcholine may affect the weighting of higher-order priors.<sup>56,57</sup> Thus, a hyperdopaminergic state (as seen in some patients with psychosis<sup>58</sup>) may increase the rate at which the speech production priors are revised, further adding to their uncertainty. D2-blockers may reduce error-induced shifts in the discourse plan, correcting the positive FTD features, but cannot improve the (nondopaminergically mediated) precision of the higher-order discourse plan per se, and the eventual gravitation toward reduced speech output (negative FTD), as seen in the long term. More broadly, neuromodulatory aberrations in the higher-level association cortex result in excitation/inhibition imbalance that underwrites precision modulation<sup>38</sup> as well as FTD<sup>59</sup> in schizophrenia. Specific brain regions, such as anterior insula,<sup>60</sup> dorsal anterior cingulate cortex<sup>61</sup> and hippocampus,<sup>62</sup> appear to play a key part in precision-weighting mechanisms; the course of FTD may relate to the volumetric<sup>63</sup> and connectivity<sup>64,65</sup> changes of these regions in schizophrenia. It is important to note that irrespective of the location of lesions, the hierarchical precision-weighting framework enables propagation of failures at 1 level to the other levels of speech generation (see Palaniyappan<sup>66</sup> for further discussion on the broader systems theory notion in psychosis).

In summary, FTD can be cast as a problem of precision modulation of higher-order priors during dyadic social encounters. This view concurs with that of Brown and Kuperberg,<sup>9</sup> who also proposed a role for imprecise priors (message-level predictions) and the resulting over-adjustment of the production plan. Our extension emphasizes a mismatch in the precision of shared priors in the speaker and listener as a crucial component of communication failure, and an inability to afford precision to discourse plans when demands arise (e.g., in a volatile social context) as a critical deficit in schizophrenia. Our model explains both positive and negative FTD and accommodates their trajectory and treatment response.

## Putting ideas to test

The active inference account of communication deficits presented here casts many elements of FTD as a 2-brain problem of social inference. As such, scientific scrutiny of this theory mandates ecologically valid experiments to build empirical evidence. Psycholinguistic studies in psychosis generally focus on lexical or sentence-level processing rather than dialogic interactions; novel insights on communication emerge when natural conversation is studied, even at a single-brain level (for example see Castellucci and colleagues<sup>67</sup>). Second-person paradigms (e.g., hyperscanning to



**Box 2. The fallacy of brain as a computer**

Throughout this editorial, we take the stance of considering the brain as a statistical engine. This is in keeping with the theoretical perspective of active inference, wherein the neuronal precision-weighting processes are described as “subpersonal,”<sup>20</sup> to distinguish them from the more conventional, affect-laden descriptions of these terms (e.g., “confidence” or “motivation” is described as precision afforded to higher-order priors such as an action policy). Nevertheless, the functional consequences implied by these terms are comparable. For instance, daily life is replete with examples where affordance of motivation is necessary for actions with highly variable consequences (e.g., going to the gym, buying lottery tickets). One objection to this “subpersonal” stance is that brains do not compute or intend; it is the person with the brain who does such computations and holds intentions. The mereological fallacy here is that we ascribe functions of the whole (person) to its parts (brains or neuronal units). Although detailed counterpoints against this objection can be found elsewhere,<sup>76</sup> it is important to ask when, if ever, can we attribute the function of a whole to one of its parts. We can say “a car moves” and “a wheel moves,” but we cannot say “a steering moves.” We can also say “I drive a car,” but we cannot convey much by saying “I drive a wheel.” In other words, for certain qualified functions, the use of intentional predicates for certain parts is agreeable. Furthermore, if we consider human intention as a socially emergent phenomenon,<sup>77</sup> it cannot be ascribed to any single part (e.g., a neuronal unit) or even a single person. In our view, these issues do not diminish the utility of the proposal made here; nevertheless, they pose a far from trivial challenge when applied to the philosophy of psychiatric practice.

capture interpersonal interactions<sup>68</sup> and brain–computer interfaces<sup>69</sup>) offer tantalizing possibilities in this regard. Indeed, emerging insights on “interbrain synchrony” of neural dynamics from healthy individuals have opened the door for second-person neuroscience in psychiatry.<sup>70</sup> Many applications for investigating FTD can emerge from the advances being made in setting up (e.g., leader–follower experiments<sup>71</sup>), recording (e.g., functional near-infrared spectroscopy) and calculating (e.g., Granger causal models) interbrain dynamics, some involving more than 2 interacting agents in a social context.<sup>24</sup>

Parameters of precision modulation can be obtained from readily interpretable behavioural readouts such as eye movements (for example see Mirza and colleagues<sup>72</sup>); such readouts can be acquired in conjunction with speech measures as a first step to test general propositions linking precision modulation and FTD. While concepts such as commitment to a discourse plan do not have direct measures, the emerging application of Natural Language Processing is already providing several indirect means to estimate parameters of relevance (see Voleti and colleagues<sup>73</sup> for a detailed review). This will enable us to acquire speech-derived measures of precision from snapshots of social discourse. Controlled experiments on certain neurophysiological phenomena related to predictive processing (e.g., electroencephalography/magnetoencephalography correlates of corollary discharge, mismatch negativity, certain evoked response potentials such as N400) can provide measures of “synaptic gain” reflecting the physiology of precision modulation.<sup>49</sup> Further, using perturb-and-measure neuromodulation techniques like transcranial direct current stimulation and focal transcranial magnetic stimulation, we can test brain-network-level specificity for these hypotheses. Such studies have been providing interesting leads to understand symptoms such as auditory verbal hallucinations in schizophrenia,<sup>74</sup> and can also

help uncover specific neurostimulation targets for FTD. The conceptual relationship between the precision of action priors and motivation raises the possibility that social incentives may nurture the physiologic markers of “synaptic gain” (representing precision) and ultimately, verbal productivity itself.<sup>75</sup> Appealing to precision dynamics also helps to explain why certain therapeutic approaches (e.g., insight-oriented therapies, group therapies) that demand one to infer and attune with another person’s priors repeatedly fail in schizophrenia. Following this line of investigation may potentially offer new therapeutic speech and language approaches in psychosis.

Though active inference has provided biologically grounded arguments to blur the boundary between “willing” and “acting” since the time of Schopenhauer, in its current edition, active inference is still a theoretical framework (Box 2). At times, this framework has been criticized as being too committed to the statistical notions of brain function, being somewhat abductive and, to some extent, dismissive of more conventional domain-specific neuropsychological concepts.<sup>78–80</sup> While its simplicity and physiologic plausibility have propelled its wide applications,<sup>81</sup> it is necessary to consider that empirical evidence is still lacking for many of its premises; for example, “synaptic gain represents precision,” “neurons maintain representational codes for various aspects of the brain’s models of the world,” “precision can be amplified at a subpersonal level,” and “compensatory changes to balance precision weights occur in pathological states” (see Keller and colleagues<sup>82</sup> for a review of evidence on neural implementation). Nevertheless, this heuristic has been immensely useful for mapping observed psychopathology to brain functions. In the case of FTD, which continues to elude neurobiological explanation, it offers a principled entry into the “dark matter” of social neuroscience.<sup>83</sup>

**Affiliations:** from the Department of Psychiatry, Schulich School of Medicine & Dentistry, Western University, London, Ont., Canada (Palaniyappan); the Robart Research Institute & Lawson Health Research Institute, London, Ont., Canada (Palaniyappan); and the InSTAR Program, Schizophrenia Clinic, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Bangalore, India (Venkatasubramanian).

**Competing interests:** L. Palaniyappan reports personal fees from Janssen Canada, Otsuka Canada, SPMM Course Limited, UK, and the Canadian Psychiatric Association; book royalties from Oxford University Press; investigator-initiated educational grants from Sunovion, Janssen Canada and Otsuka Canada; and travel support from Boehringer Ingelheim and Magstim Limited, outside the submitted work. No other competing interests were declared.

**Funding:** L. Palaniyappan’s work is supported by the Canadian Institute of Health Research (foundation grant 375104), Bucke Family Fund and the Tanna Schulich Endowment Chair. G. Venkatasubramanian’s work is supported by the Department of Biotechnology (DBT) — Wellcome Trust India Alliance (IA/CRC/19/1/610005) and the Department of Biotechnology, Government of India (BT/HRD-NBA-NWB/38/2019-20(6)).

**Content licence:** This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

1. Schopenhauer A, Payne EFJ. *The world as will and representation*. Dover Publications; 1966.
2. Ike KGO, de Boer SF, Buwalda B, et al. Social withdrawal: an initially adaptive behavior that becomes maladaptive when expressed excessively. *Neurosci Biobehav Rev* 2020;116:251-67.
3. Kingsbury L, Huang S, Wang J, et al. Correlated neural activity and encoding of behavior across brains of socially interacting animals. *Cell* 2019;178:429-446.e16.
4. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 2004;27:712-9.
5. Adams RA, Shipp S, Friston KJ. Predictions not commands: active inference in the motor system. *Brain Struct Funct* 2013;218:611-43.
6. Friston KJ, Daunizeau J, Kiebel SJ. Reinforcement learning or active inference? *PLoS One* 2009;4:e6421.
7. Petzschner FH, Weber LAE, Gard T, et al. Computational psychosomatics and computational psychiatry: toward a joint framework for differential diagnosis. *Biol Psychiatry* 2017;82:421-30.
8. Limongi R, Jeon P, Mackinley M, et al. Glutamate and dysconnection in the salience network: neurochemical, effective connectivity, and computational evidence in schizophrenia. *Biol Psychiatry* 2020;88:273-81.
9. Brown M, Kuperberg GR. A hierarchical generative framework of language processing: linking language perception, interpretation, and production abnormalities in schizophrenia. *Front Hum Neurosci* 2015;9:643.
10. Lecaigard F, Bertrand O, Caclin A, et al. Neurocomputational underpinnings of expected surprise. *J Neurosci* 2022;42:474-86.
11. Yon D, Frith CD. Precision and the Bayesian brain. *Curr Biol* 2021;31:R1026-32.
12. Clark A. Predictions, precision, and agentive attention. *Conscious Cogn* 2017;56:115-9.
13. Parr T, Benrimoh DA, Vincent P, et al. Precision and false perceptual inference. *Front Integr Neurosci* 2018;12:39.
14. Pezzulo G, Rigoli F, Friston KJ. Hierarchical active inference: a theory of motivated control. *Trends Cogn Sci* 2018;22:294-306.
15. Faivre N, Roger M, Pereira M, et al. Confidence in visual motion discrimination is preserved in individuals with schizophrenia. *J Psychiatry Neurosci* 2021;46:E65-73.
16. Smith R, Kirlic N, Stewart JL, et al. Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: a computational modelling approach. *J Psychiatry Neurosci* 2021;46:E74-87.
17. Powers AR, Mathys C, Corlett PR. Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* 2017;357:596-600.
18. Paulus MP, Feinstein JS, Khalsa SS. An active inference approach to interoceptive psychopathology. *Annu Rev Clin Psychol* 2019;15:97-122.
19. Barrett LF, Quigley KS, Hamilton P. An active inference theory of allostasis and interoception in depression. *Philos Trans R Soc Lond B Biol Sci* 2016;371:20160011.
20. Vasil J, Badcock PB, Constant A, et al. A world unto itself: human communication as active inference. *Front Psychol* 2020;11:417.
21. Friston KJ, Parr T, Yufik Y, et al. Generative models, linguistic communication and active inference. *Neurosci Biobehav Rev* 2020;118:42-64.
22. Friston K, Frith C. A duet for one. *Conscious Cogn* 2015;36:390-405.
23. Friston KJ, Frith CD. Active inference, communication and hermeneutics. *Cortex* 2015;68:129-43.
24. Kingsbury L, Hong W. A multi-brain framework for social interaction. *Trends Neurosci* 2020;43:651-66.
25. Murphy E, Holmes E, Friston K. Natural language syntax complies with the free-energy principle. (2021) doi:10.31234/osf.io/r3fxc.
26. Veissière SPL, Constant A, Ramstead MJD, et al. Thinking through other minds: a variational approach to cognition and culture. *Behav Brain Sci* 2019;43:e90.
27. Parr T, Limanowski J, Rawji V, et al. The computational neurology of movement under active inference. *Brain* 2021;144:1799-1818.
28. Chaika E. Thought disorder or speech disorder in schizophrenia? *Schizophr Bull* 1982;8:587-94.
29. Hinzen W. What is Un-Cartesian linguistics? *Biolinguistics* 2014;8:226-57.
30. Hinzen W. Reference across pathologies: a new linguistic lens on disorders of thought. *Theor Linguist* 2017;43:169-232.
31. Wible CG. Schizophrenia as a disorder of social communication. *Schizophr Res Treatment* 2012;2012:920485.
32. Docherty NM, DeRosa M, Andreassen NC. Communication disturbances in schizophrenia and mania. *Arch Gen Psychiatry* 1996;53:358-64.
33. Hinzen W, Rosselló J. The linguistics of schizophrenia: thought disturbance as language pathology across positive symptoms. *Front Psychol* 2015;6:971.
34. Cutting J. A general overview. *J Ital Soc Psychopathol* 2016;22:48-54.
35. Parola A, Gabbatore I, Berardinelli L, et al. Multimodal assessment of communicative-pragmatic features in schizophrenia: a machine learning approach. *NPJ Schizophr* 2021;7:28.
36. Adamczyk P, Daren A, Sulecka A, et al. Do better communication skills promote sheltered employment in schizophrenia? *Schizophr Res* 2016;176:331-9.
37. Agostoni G, Bambini V, Bechi M, et al. Communicative-pragmatic abilities mediate the relationship between cognition and daily functioning in schizophrenia. *Neuropsychology* 2021;35:42-56.
38. Sterzer P, Adams RA, Fletcher P, et al. The predictive coding account of psychosis. *Biol Psychiatry* 2018;84:634-43.
39. Horga G, Abi-Dargham A. An integrative framework for perceptual disturbances in psychosis. *Nat Rev Neurosci* 2019;20:763-778.
40. Leroy F, Pezard L, Nandrino J-L, et al. Dynamical quantification of schizophrenic speech. *Psychiatry Res* 2005;133:159-71.
41. Press C, Kok P, Yon D. The perceptual prediction paradox. *Trends Cogn Sci* 2020;24:13-24.
42. Gozé T, Naudin J. Discussing Rümke's "Praecox feeling" from the clinician's experience of schizophrenic contact. *Revista Psicopatologia Fenomenologica Contemporanea* 2017;6:112-23.
43. van Neerven T, Bos DJ, van Haren NE. Deficiencies in theory of mind in patients with schizophrenia, bipolar disorder, and major depressive disorder: a systematic review of secondary literature. *Neurosci Biobehav Rev* 2021;120:249-61.
44. Dwyer K, David AS, McCarthy R, et al. Linguistic alignment and theory of mind impairments in schizophrenia patients' dialogic interactions. *Psychol Med* 2020;50:2194-202.
45. Sandford DA, Elzinga RH. Patient interactions in four psychiatric wards. *Acta Psychiatr Scand* 1990;82:125-9.
46. Jeganathan J, Breakspear M. An active inference perspective on the negative symptoms of schizophrenia. *Lancet Psychiatry* 2021;8:732-8.
47. Gherman S, Philiastides MG. Neural representations of confidence emerge from the process of decision formation during perceptual choices. *Neuroimage* 2015;106:134-43.
48. Sherman BE, Graves KN, Turk-Browne NB. The prevalence and importance of statistical learning in human cognition and behavior. *Curr Opin Behav Sci* 2020;32:15-20.
49. Adams RA, Pinotsis D, Tsirlis K, et al. Computational modeling of electroencephalography and functional magnetic resonance imaging paradigms indicates a consistent loss of pyramidal cell synaptic gain in schizophrenia. *Biol Psychiatry* 2022;91:202-15.
50. Nakazawa K, Jeevakumar V, Nakao K. Spatial and temporal boundaries of NMDA receptor hypofunction leading to schizophrenia. *NPJ Schizophr* 2017;3:7.
51. Deserno L, Boehme R, Mathys C, et al. Volatility estimates increase choice switching and relate to prefrontal activity in schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2020;5:173-83.
52. Wilkinson S, Dodgson G, Meares K. Predictive processing and the varieties of psychological trauma. *Front Psychol* 2017;8:1840.

53. Haarsma J, Knolle F, Griffin JD, et al. Influence of prior beliefs on perception in early psychosis: effects of illness stage and hierarchical level of belief. *J Abnorm Psychol* 2020;129:581-98.
54. Lawson RP, Bisby J, Nord CL, et al. The computational, pharmacological, and physiological determinants of sensory learning under uncertainty. *Curr Biol* 2021;31:163-172.e4.
55. Valdés-Baizabal C, Carbajal GV, Pérez-González D, et al. Dopamine modulates subcortical responses to surprising sounds. *PLoS Biol* 2020;18:e3000744.
56. Iglesias S, Kasper L, Harrison SJ, et al. Cholinergic and dopaminergic effects on prediction error and uncertainty responses during sensory associative learning. *Neuroimage* 2021;226:117590.
57. Iglesias S, Mathys C, Brodersen KH, et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* 2013;80:519-30.
58. Brugger SP, Angelescu I, Abi-Dargham A, et al. Heterogeneity of striatal dopamine function in schizophrenia: meta-analysis of variance. *Biol Psychiatry* 2020;87:215-24.
59. Palaniyappan L. Dissecting the neurobiology of linguistic disorganisation and impoverishment in schizophrenia. *Semin Cell Dev Biol* 2021 [Epub ahead of print]. doi: 10.1016/j.semdb.2021.08.015.
60. Corlett PR, Mollick JA, Kober H. Meta-analysis of human prediction error for incentives, perception, cognition, and action. *Neuropsychopharmacology* 2022 [Epub ahead of print]. doi: 10.1038/s41386-021-01264-3.
61. Haarsma J, Fletcher PC, Griffin JD, et al. Precision weighting of cortical unsigned prediction error signals benefits learning, is mediated by dopamine, and is impaired in psychosis. *Mol Psychiatry* 2021;26:5320-33.
62. Barron HC, Aukstulewicz R, Friston K. Prediction and memory: a predictive coding account. *Prog Neurobiol* 2020;192:101821.
63. Li M, Li X, Das TK, et al. Prognostic utility of multivariate morphometry in schizophrenia. *Front Psychiatry* 2019;10:245.
64. Pan Y, Dempster K, Jeon P, et al. Acute conceptual disorganization in untreated first-episode psychosis: a combined magnetic resonance spectroscopy and diffusion imaging study of the cingulum. *J Psychiatry Neurosci* 2021;46:E337-46.
65. Overbeek G, Gawne TJ, Reid MA, et al. A multimodal neuroimaging study investigating resting-state connectivity, glutamate and GABA at 7 T in first-episode psychosis. *J Psychiatry Neurosci* 2021;46:E702-10.
66. Palaniyappan L. Inefficient neural system stabilization: a theory of spontaneous resolutions and recurrent relapses in psychosis. *J Psychiatry Neurosci* 2019;44:367-383.
67. Castellucci GA, Kovach CK, Howard MA, et al. A speech planning network for interactive language use. *Nature* 2022;1-6 10.1038/s41586-021-04270-z.
68. Jiang J, Zheng L, Lu C. A hierarchical model for interpersonal verbal communication. *Soc Cogn Affect Neurosci* 2021;16:246-55.
69. Chame HF, Ahmadi A, Tani J. A hybrid human-neurorobotics approach to primary intersubjectivity via active inference. *Front Psychol* 2020;11:584869.
70. Kelsen BA, Sumich A, Kasabov N, et al. What has social neuroscience learned from hyperscanning studies of spoken communication? A systematic review. *Neurosci Biobehav Rev* 2020 [Epub ahead of print]. doi: 10.1016/j.neubiorev.2020.09.008.
71. Jiang J, et al. Leader emergence through interpersonal neural synchronization. *Proc Natl Acad Sci U S A* 2015;112:4274-9.
72. Mirza MB, Cullen M, Parr T, et al. Contextual perception under active inference. *Sci Rep* 2021;11:16223.
73. Voleti R, Liss JM, Berisha V. A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE J Sel Top Signal Process* 2020;14:282-98.
74. Bose A, Nawani H, Agarwal SM, et al. Effect of fronto-temporal transcranial direct current stimulation on corollary discharge in schizophrenia: a randomized, double-blind, sham-controlled mediation analysis study. *Schizophr Res* 2019;204:411-2.
75. Joyal M, Bonneau A, Fecteau S. Speech and language therapies to improve pragmatics and discourse skills in patients with schizophrenia. *Psychiatry Res* 2016;240:88-95.
76. Pöyhönen S. Intentional concepts in cognitive neuroscience. *Philos Explor* 2014;17:93-109.
77. Gibbs, RW Jr. Intentions as emergent products of social interactions. In: Malle BF, Moses LJ, Baldwin DA (editors). *Intentions and intentionality: Foundations of social cognition*. The MIT Press; 2001;105-122.
78. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 2013;36:181-204.
79. Kogo N, Trengove C. Is predictive coding theory articulated enough to be testable? *Front Comput Neurosci* 2015;9:111.
80. Williams D. Predictive coding and thought. *Synthese* 2020;197:1749-75.
81. Gadsby S, Hohwy J. Why use predictive processing to explain psychopathology? The case of anorexia nervosa. *Research Gate* 2019 Apr. 13 [preprint]. doi:10.31234/osf.io/y46z5 (accessed 2022 Jan. 26).
82. Keller GB, Mrsic-Flogel TD. Predictive processing: a canonical cortical computation. *Neuron* 2018;100:424-35.
83. Przyrembel M, Smallwood J, Pauen M, et al. Illuminating the dark matter of social neuroscience: considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Front Hum Neurosci* 2012;6:190.