DOCUMENT SUMMARY

This expert review by Karl Friston presents a comprehensive framework for "strong" computational psychiatry, which seeks to explain psychopathology through underlying pathophysiology. The central argument is that the brain operates as an inferential organ, constantly creating and testing predictive models of the world, a process termed "active inference." From this perspective, psychopathology is fundamentally "false inference," arising from "synaptopathies" (dysfunctional synapses) that impair the brain's ability to precisely modulate the balance between sensory evidence (bottom-up) and prior beliefs (top-down). This paper provides a core scientific foundation for Enlitens by arguing against simplistic, descriptive models (like standardized testing) and advocating for individualized approaches that aim to understand each person's unique generative model, a process it calls "computational phenotyping."

FILENAME

Friston_K_2022_Computational_Psychiatry_Synapses_to_Sentience_framework_for_false_inference_and_precision.md

METADATA

- **Primary Category**: NEURODIVERSITY
- **Document Type**: expert_review
- **Relevance**: Core
- **Key Topics**: computational_psychiatry, false_inference, predictive_processing, neurodiversity, assessment_critique, synaptopathy, precision, computational_phenotyping
- **Tags**: #active_inference, #Bayesian_brain, #false_inference, #precision, #synaptopathy, #dysconnection, #assessment, #schizophrenia, #autism, #trauma, #computational_phenotyping

CRITICAL QUOTES FOR ENLITENS

- "The narrative starts with the simple premise that if psychology— read as belief updating in the brain—can be cast as a computational process of inference, it follows that psychopathology just is false inference."
- "The common theme that distinguishes weak from strong approaches is reference to a generative model: namely, a hypothesis about causal mechanisms, specified formally as a joint probability distribution over some (unobservable, latent or hidden) causes and their (observable, measurable) consequences."

- "In short, many psychiatric, neurological and neurodegenerative disorders could be described as arising from pernicious synaptopathies that lead to a functional disintegration of neuronal message passing in cortical and subcortical hierarchies."
- "...many psychiatric and neurological disorders can be framed in terms of a synaptopathy that confounds the encoding of uncertainty or precision in a (Bayesian) brain..."
- "Since that time, it has become difficult to find a psychiatric condition that has not been considered through the lens of aberrant precision, in one form or another [21, 135, 137–161]."
- "It means that any given patient can be fully characterised by her prior beliefs under ideal (active Bayesian inference) observer assumptions."
- "This motivates the difficult game of inferring what generative model this subject is using, based purely upon their behavioural or neuronal responses [204–206]."
- "One obvious advantage of being able to create a 'digital twin' of a patient in silico, is the ability to perform synthetic lesion studies and psychopharmacology, to test various hypotheses about the effect of therapeutic interventions."

KEY STATISTICS & EVIDENCE

- "For example, a recent genome-wide association study [122] of 76,755 individuals with schizophrenia highlights the importance of GRIN2A (which encodes a subunit of the NMDA receptor), while a recent meta-analysis of whole exomes [123] emphasises the role of GRIN2A, as well as GRIA3 (which encodes a subunit of the AMPA receptor)."
- A study on schizophrenia concluded that EEG responses in three classic paradigms are all attributable to "greater self-inhibition in pyramidal cells" and that psychotic symptoms relate to "disinhibition in neural circuits". These findings support the hypothesis that "a primary loss of synaptic gain on pyramidal cells is then compensated by interneuron downregulation (rather than the converse)".
- In patients with frontotemporal lobar degeneration, it was shown that "the phasic inhibition of deep cortico-cortical pyramidal neurons following tiagabine, but not placebo, was a function of GABA concentration" [199]."

METHODOLOGY DESCRIPTIONS

"Strong" vs. "Weak" Computational Approaches

The paper distinguishes between two types of computational methods, arguing for the superiority of "strong," theory-led approaches.

- **Weak Approaches:** These are descriptive or theory-free, such as behaviorism, analyzing correlation patterns in fMRI (functional connectivity), or using 'big data' and deep learning without a mechanistic model. These methods lack explanatory power. For example, "the correlation of a source with itself tells you nothing (it is always one)".
- **Strong Approaches:** These are hypothesis-led and rest on a "generative model"—a formal hypothesis about the causal mechanisms that produce observable data. This allows for mechanistic interpretation and explanation. Examples include active inference,

effective connectivity analysis, and generative adversarial networks. The paper advocates for this approach as it allows for comparing the evidence for different models.

Computational Phenotyping and Nosology

This is presented as a "strong" approach to characterizing individual patients, moving beyond simple classification.

- **Core Principle:** The complete class theorem suggests that for any behavior, there exists a set of prior beliefs that would make that behavior Bayes optimal. This implies "that any given patient can be fully characterised by her prior beliefs under ideal (active Bayesian inference) observer assumptions".
- **Method:** The goal is to infer the generative model a subject is using based on their behavioral or neuronal responses. This is sometimes called "generative embedding". This approach can be more efficient for classifying or stratifying patients than using their raw responses.
- **Computational Nosology:** This extends the idea to model the slow, dynamic processes of pathophysiology and psychopathology over time. Conventional diagnoses are not used to stratify patients but as "data features that supply evidence for or against models of the mapping between pathophysiology and psychopathology that are hidden from direct observation".

THEORETICAL FRAMEWORKS

Psychopathology as False Inference

"The narrative starts with the simple premise that if psychology— read as belief updating in the brain—can be cast as a computational process of inference, it follows that psychopathology just is false inference. False inference is meant in the usual sense of false positives (i.e., type I errors); namely, inferring something is there when it is not. Cardinal examples here include hallucinations, delusions and other features of reality distortion seen in psychosis. False negatives (i.e., type II errors) mean inferring something is not there when it is; for example, dissociative disorders, neglect syndromes, derealisation phenomena, et cetera. Indeed, when one thinks about psychiatric and neurological disorders, most can be framed as false inference: ranging from dysmorphophobia in eating disorders, through to delusional systems in paranoid schizophrenia; from phobias through to Parkinson's disease."

The Importance of Precision Engineering

Precision is the confidence or certainty afforded to prediction errors, which is biologically encoded by the gain or excitability of neuronal populations. This mechanism is central to attention.

"Affording certain prediction errors greater precision increases their influence on belief updating and has all the hallmarks of attentional selection [80–85]. Physiologically, this simply entails an increase in the excitability or postsynaptic gain of neuronal populations broadcasting prediction errors. ... One crucial aspect of this precision engineering is that it underwrites our ability to filter out—or ignore—certain prediction errors when they are deemed imprecise. A key example is

the attenuation of sensory prediction errors that report the consequences of movement [93–98). If we could not ignore the proprioceptive and somatosensory afferents—supplying evidence that we are not moving—then any beliefs about intended or predicted movement would be revised immediately, and we would not be able to initiate movement." This failure of "sensory attenuation" is a key theme across many conditions.

Synaptopathy and Dysconnection as the Root of False Inference

The paper argues that false inference stems from biological issues at the synaptic level. "In short, one could neatly summarise the pathophysiology of many psychiatric and neurological conditions in terms of one or more forms of synaptopathy [108]. Synaptopathy is taken to mean any failure of synaptic function due to a variety of pathological mechanisms (i.e., formation, structure, metabolism, etc.). ... A coarse-grained overview of the synaptic theories of schizophrenia suggests that the synaptopathy in question is of a neuromodulatory sort; implicating classical neuromodulatory (ascending) neurotransmitter systems [110–113), GABAergic neurotransmission and NMDA receptors [113–116]...".

This leads to the central thesis: "the psychopathology (i.e., false inference) characteristic of psychiatric disorders may be attributable to aberrant precision control, which inherits from synaptopathies that confound neuromodulation."

POPULATION-SPECIFIC FINDINGS

The paper frames various conditions through the lens of aberrant precision and sensory attenuation:

- **A Crosscutting Theme:** "A crosscutting theme in many of these accounts is a putative failure of sensory attenuation... A failure of sensory attenuation implies an imbalance between the precision afforded sensory prediction errors and the precision of prediction errors deeper in neuronal hierarchies that mediate or maintain prior beliefs... The ensuing imbalance is often read as a loss of precision or confidence in prior beliefs, relative to the sensory evidence at hand."
- **Autism:** "In autism, this imbalance is consistent with a failure of sensory attenuation and an inability to ignore the sensorium, which may or may not be associated with the neuromodulatory role of oxytocin [135, 166–170]."
- **Schizophrenia:** "...a failure of sensory attenuation provides an apt explanation for resistance to illusions (that normally depend upon precise prior beliefs) and failures to elicit mismatch oddball responses (because everything is surprising). Some people then interpret delusional ideation as the brain's attempt to make sense of unattenuated prediction errors... However, hallucinatory phenomena require a slightly more delicate argument; usually along the lines of a compensatory increase in prior precision that could manifest as a form of paradoxical lesion. In other words, in attempt to override precise sensory prediction errors, higher levels learn to ignore sensory evidence and—sequestered from the sensorium—elaborate false percepts."
- **Other Conditions:** The framework is applied broadly, "ranging from failures of sensory attenuation in autonomic and interoceptive inference, through depression and emotional processing [82, 140, 149, 174], to functional medical symptoms and dissociative phenomena [137, 164]."

PRACTICAL APPLICATIONS

In Vivo Assays of Synaptic Efficacy

"...this calls for in vivo and ex vivo assays of synaptic efficacy [178, 179] that can be deployed across scale; i.e., from molecular biology through to patients in the clinic (or bedside)." Dynamic Causal Modelling (DCM) is presented as the primary "strong" approach for this in imaging neuroscience. It models neuronal dynamics to estimate intrinsic and extrinsic connectivity, allowing researchers to assess the evidence for changes in synaptic efficacy.

Computational Neuropsychology (The "Digital Twin")

"A final part of the narrative is that a generative model of pathophysiology, and ensuing psychopathology, allows one to perform in silico or synthetic experiments [212–214]. These entail optimising the parameters of a model of a particular subject or cohort, and seeing what would happen if one increased or decreased certain synaptic efficacies, e.g., [197]. ... One obvious advantage of being able to create a 'digital twin' of a patient in silico, is the ability to perform synthetic lesion studies and psychopharmacology, to test various hypotheses about the effect of therapeutic interventions."