Scientific Thesis on the Model of Human States over Time (MHST)

A critical analysis of the theoretical possibilities and empirical limits of the MHST framework and its formalization through Predictive Processing

Author	Unknown Author
Date	September 2026
Version	Final
Discipline	Cognitive Science / Neuroinformatics
Sources	35 cited sources

Table of Contents

Abstract

The Model of Human States over Time (MHST) describes subjective perception as a temporally dynamic process in which incoming information interacts with an internal cognitive state vector $X(t) = [T(t), B(t), A(t), SC(t), CR(t)]$. The formalization through the Predictive Processing approach (Friston, 2010[5]; Clark, 2013[6]; Hohwy, 2013[9]) replaces the abstract transformation functions F and G with variational free energy minimization and gradient descent on precision-weighted prediction errors. This thesis argues that the MHST model provides a theoretically coherent framework that generates three central predictions: (1) state dependence of perception, (2) path dependence (cognitive hysteresis), and (3) multiscale dynamics of cognitive structures. These predictions are consistent with established empirical findings from cognitive psychology, including anchoring effects (Tversky & Kahneman, 1974[1]), confirmation bias (Wason, 1960[2]; Nickerson, 1998[3]), and attitude polarization (Lord, Ross & Lepper, 1979[4]), as well as with neurophysiological markers such as mismatch negativity (Wacongne et al., 2011[11]; Bendixen et al., 2012[12]) and Bayesian learning rates (Behrens et al., 2007[10]). **However, the model is currently a conceptual proposal without empirical validation.** Its five variables are theoretical constructs without established operationalization in the sense of construct validity (Cronbach & Meehl, 1955[16]), its parameter values are illustrative anchor points from foreign paradigms, and the five-variable architecture neglects important additional factors such as emotion, attention, memory, and social context. Empirical validation would require a multi-year research program, beginning with the operationalization of two to three variables in a controlled longitudinal design.

Keywords: Subjective perception, Predictive Processing, Free-Energy Principle, Active Inference, state dependence, path dependence, cognitive hysteresis, construct validity, Computational Psychiatry, Bayesian inference, hierarchical processing

1. Introduction

The question of how the human brain constructs a coherent subjective experience from incomplete and ambiguous sensory signals is one of the central problems of cognitive science. Perception is not a passive imprint of the external world, but an active constructive process in which prior knowledge, expectations, and beliefs play a decisive role (Friston, 2010[5]; Clark, 2013[6]). Helmholtz (1867) already postulated that perception is based on *unconscious inference* — an idea that has been further developed in modern form as *Predictive Processing* or *Active Inference* (Hohwy, 2013[9]; Seth, 2021[15]).

The Model of Human States over Time (MHST) is a conceptual framework that formally describes this constructive process. It represents the cognitive state of a person as the vector X(t) = [T(t), B(t), A(t), SC(t), CR(t)], comprising thoughts (T), beliefs (B), attitudes (A), unconscious processes (SC), and conscious reality (CR). Incoming information I(t) is transformed by this state and produces a conscious representation, which in turn modifies the cognitive state when processing future information. The extended synthesis replaces the abstract functions F and G with the mathematically explicit mechanics of Predictive Processing, whereby SC(t) is operationalized as a precision matrix and CR(t) is defined as the minimum of variational free energy (Friston, 2010[5]).

This thesis systematically examines which scientific claims are possible with the MHST model and where its epistemic limits lie. The work pursues a threefold goal: (1) to examine the theoretical coherence of the model, (2) to assess its empirical testability, and (3) to identify its conceptual limitations.

2. Theoretical Background

2.1 The Free-Energy Principle

The Free-Energy Principle (FEP) postulates that biological systems act as minimizers of variational free energy — an information-theoretic measure that provides an upper bound on the surprise (negative log-likelihood) of sensory data (Friston, 2010[5]). Mathematically, free energy decomposes into a complexity term (which measures the divergence between the internal estimate q(theta) and the true posterior) and an accuracy term (which measures the expected log-likelihood of sensory data under the estimate). The brain minimizes this value by adjusting its internal states, mathematically formulated as gradient descent (Bastos et al., 2012[20]; Lee & Mumford, 2003[21]).

2.2 Hierarchical Predictive Coding

Predictive Coding goes beyond the FEP by postulating a specific neural architecture: a hierarchical system in which each level sends predictions to the level below and passes prediction errors upward (Rao & Ballard, 1999[22]; Clark, 2013[6]). The updating of internal states occurs proportional to the precision-weighted prediction error, whereby precision is defined as the inverse variance of the error distribution and is modulated by attention (Feldman & Friston, 2010[13]). This architecture has been linked to the canonical microcircuits of the neocortex, with prediction errors in superficial pyramidal cells of layers II/III and predictions in deep pyramidal cells of layers V/VI (Bastos et al., 2012[20]).

2.3 Construct Validity and Psychological Measurement

Operationalizing psychological constructs requires more than assigning a name. Cronbach and Meehl (1955[16]) introduced the concept of *construct validity* and argued that a construct must be validated through a *nomological network* — a web of theoretical relationships supported by empirical data. This is particularly relevant for the MHST model, whose five variables are theoretical constructs without established measurement models. Criterion validation requires convergent validity (the measurement instrument correlates with other measurement instruments of the same construct) and discriminant validity (the measurement instrument does not correlate with measurement instruments of different constructs).

3. Summary of the MHST Model

3.1 The Five Core Variables and Their Time Hierarchy

The cognitive state of a person at time t is represented by the vector:

Equation 1 — The MHST state vector

$$X(t) = [\ T(t), B(t), A(t), SC(t), CR(t)\ ]^T \text{ in } R^d$$

The model postulates a strict time-constant hierarchy:

Equation 2 — Temporal scale separation

$$tau(T) << tau(A) << tau(B) << tau(SC)$$

This hierarchy states that thoughts change on the order of milliseconds to seconds, attitudes over hours to weeks, beliefs over months to years, and unconscious structures over the lifespan. This prediction is qualitatively supported by neurophysiological data: mismatch negativity (MMN), an ERP marker for automatic prediction-error detection, peaks at approximately 150–250 ms post-stimulus (Wacongne et al., 2011[11]; Bendixen et al., 2012[12]), while Bayesian learning rates for belief updating are around 0.3–0.7 and operate across multiple trials (Behrens et al., 2007[10]).

Variable	Construct	Time const.	Role in the system
T(t)	Thoughts	ms to s	Fleeting representations in working memory.
B(t)	Beliefs	Days to yrs	Propositional world models; interpretation structures.
A(t)	Attitudes	Hrs to wks	Valenced orientations; modulate significance.
SC(t)	Unconscious	Lifespan	Operationalized as precision matrix Pi(t).
CR(t)	Conscious Reality	Real time	Converged posterior after free-energy minimization.

3.2 The Central Functions and Their Operationalization

The model is described by two functions. The **perceptual transformation F** maps incoming information and the internal state onto conscious experience:

Equation 3 — Perceptual transformation (abstract form)

$$CR(t) = F[\ I(t), T(t), B(t), A(t), SC(t)\ ]$$

The **state transition function G** governs the continuous updating of the state vector:

Equation 4 — State transition equation (continuous)

$$dX(t)/dt = G(X(t), I(t), CR(t))$$

Through formalization with Predictive Processing, these abstract functions are replaced by concrete mathematical mechanisms. F is operationalized as minimization of variational free energy (Friston, 2010[5]):

Equation 5 — Variational free energy (general form)

$$F = E\_q[\ln q(mu) - \ln p(s, mu)]$$

Equation 6 — Quadratic free energy (Laplace approximation)

$$\sim 1/2 * Sum\_l\ Pi\_l(t) * eps\_l(t)^2 + const.$$

Equation 7 — Conscious reality as free-energy minimum

$$CR(t) = argmin\_mu\ F$$

where eps_l(t) is the prediction error at level l and Pi_l(t) is the precision weighting. G is operationalized as gradient descent on this free energy (Bastos et al., 2012[20]; Lee & Mumford, 2003[21]):

Equation 8 — Gradient descent (continuous)

$$dmu\_l/dt = -kappa\_l [Pi\_l(t) * eps\_l(t) - (dg\_l/dmu\_l)^T * Pi\_\{l-1\}(t) * eps\_\{l-1\}(t)]$$

Equation 9 — Euler discretization for simulation

$$mu\_\{l, t+1\} = mu\_\{l, t\} - dt * kappa\_l [Pi\_l * eps\_\{l, t\} - (dg\_l/dmu\_l)^T * Pi\_\{l-1\} * eps\_\{l-1, t\}]$$

The precision matrix Pi_l(t) replaces SC(t) and is defined as the inverse covariance matrix of the prediction errors; it is modulated by attention and neuromodulators such as acetylcholine and noradrenaline (Feldman & Friston, 2010[13]; Yu & Dayan, 2005[23]):

Equation 10 — Precision as inverse variance

$$Pi\_l(t) = Sigma\_l(t)^{-1} = 1/sigma\_l^2(t)$$

The prediction errors at each hierarchical level are computed as:

Equation 11 — Hierarchical prediction error

$$eps\_l(t) = mu\_l(t) - g\_l(mu\_\{l+1\}(t))$$

Equation 12 — Sensory prediction error

$$eps\_1(t) = s(t) - g\_1(mu\_2(t))$$

where g_l is the generative function that maps level l+1 onto level l, and s(t) is the sensory input. The MHST-to-PP mapping is summarized below:

MHST Construct	PP Equivalent	Mathematical Operationalization
Information I(t)	Sensory stream s(t)	I(t) = s(t)
Beliefs B(t)	Hierarchical priors	B(t) = {mu_2, ..., mu_n}
Attitudes A(t)	Valenced priors	Prior preferences over states
Thoughts T(t)	Top-down predictions	T(t) = g(mu_{l+1})
Unconscious SC(t)	Precision matrix Pi	SC(t) = Sigma^(-1)(t)
Conscious Reality CR(t)	Converged posterior mu*	CR(t) = argmin F

3.3 Path Dependence and Cognitive Hysteresis

A central theoretical feature of the MHST model is its explicit rejection of memoryless processing. When a person first processes information I_1 and then I_2, the second input is processed through the already modified state X_1, not through the original state X_0:

Equation 13 — First conscious representation

CR_1 = F(X_0, I_1)

Equation 14 — State update after the first event

X_1 = X_0 + integral_{t0 to t1} G(X(t), I_1, CR(t)) dt

Equation 15 — Second representation depends on the first

CR_2 = F(X_1, I_2) = F(G(X_0, I_1, CR_1), I_2)

Since F and G are nonlinear operators, reversing the order of information yields a fundamentally different conscious outcome:

Equation 16 — Non-commutativity (cognitive hysteresis)

CR_2' = F(I_1, X_0 + integral G(X(t), I_2, F(I_2, X_0)) dt) != CR_2

The model calls this phenomenon **cognitive hysteresis**. This prediction is consistent with classic findings on primacy effects in impression formation (Asch, 1946[24]) and order effects in judgment (Hogarth & Einhorn, 1992[25]).

3.4 The Five Feedback Loops

The model postulates five feedback loops through which conscious experience acts back on the state components. The first loop is the sensory drive (I -> CR). The remaining four loops have explicit update equations:

Equation 17 — Loop 2: thought induction (CR -> T)

T(t+dt) = f_2(CR(t))

Equation 18 — Loop 3: belief consolidation (CR -> B)

B(t+dt) = B(t) + eta_B * h(CR(t))

Equation 19 — Loop 4: attitude adjustment (CR -> A)

A(t+dt) = A(t) + eta_A * g(CR(t))

Equation 20 — Loop 5: unconscious weighting (CR -> SC)

$$SC(t+dt) = SC(t) + \int CR(\tau)\, k(t-\tau)\, d\tau$$

where eta_B and eta_A are learning rates and k(t-tau) is a kernel function that governs the temporal integration of conscious experience into unconscious structures. This recurrent architecture is consistent with the concept of *Active Inference* (Friston et al., 2010[26]; Friston, 2017[27]), in which an agent actively shapes its environment in order to minimize prediction errors.

3.5 Formal Mathematical Equations

In addition to the equations above, the MHST model proposes several further formalizations. The perceptual transformation can be decomposed into four sub-functions:

Equation 21 — Decomposed perceptual transformation

$$CR(t) = \mathrm{INTEGRATE}\,[\, F_T(I(t), T(t)),\ F_B(I(t), B(t)),\ F_A(I(t), A(t)),\ F_{SC}(I(t), SC(t)) \,]$$

A simplified linear approximation is offered, with the explicit caveat that a realistic system would require nonlinear interaction terms:

Equation 22 — Linear approximation (first order)

$$CR(t) = w_I \cdot I(t) + w_T \cdot T(t) + w_B \cdot B(t) + w_A \cdot A(t) + w_{SC} \cdot SC(t) + \varepsilon(t)$$

For future empirical work, a discrete-time state-space form is proposed:

Equation 23 — Discrete state transition

$$X_{t+1} = A \cdot X_t + B \cdot I_t + U_t$$

Equation 24 — Discrete observation equation

$$CR_t = C \cdot X_t + D \cdot I_t + \varepsilon_t$$

A spatio-temporal extension adds spatial coordinates s in $\mathbb{R}^3$:

Equation 25 — Spatio-temporal field dynamics

$$dX(s,t)/dt = D \cdot \nabla^2 X(s,t) + G(\, X(s,t), I(s,t), CR(s,t) \,)$$

Finally, at convergence, conscious reality can be written as the argmin of the precision-weighted squared prediction errors across all hierarchical levels:

Equation 26 — Conscious reality as precision-weighted error minimum

$$CR(t) = \arg\min_{\mu} \left(1/2 \cdot \sum_l \Pi_l(t)\, [\, \mu_l - g_l(\mu_{l+1}) \,]^2 \right)$$

4. Scientific Thesis

Central thesis: The MHST model represents a theoretically coherent and mathematically specifiable conceptual framework that describes the construction of subjective perception as an interaction of incoming information with a temporally evolving cognitive state. Its three central predictions — state dependence, path dependence, and multiscale dynamics — are in principle testable and consistent with the established empirical literature. **However, the model is currently a conceptual proposal, not an empirically validated measurement model.** Its variables are theoretical constructs without established operationalization in the sense of construct validity (Cronbach & Meehl, 1955[16]), its parameter values are illustrative anchor points from foreign paradigms, and it neglects important additional variables. Empirical validation would require a multi-year research program.

4.1 Formal Hypotheses

Five formally testable hypotheses can be derived from the MHST model:

> **H1 (State Dependence):** Identical information, presented to people with different prior cognitive states, should produce measurably different interpretations. Consistent with anchoring effects (Tversky & Kahneman, 1974[1]) and confirmation bias (Wason, 1960[2]; Nickerson, 1998[3]).

> **H2 (Path Dependence):** The interpretation of I_2 should partly depend on prior exposure to I_1 if I_1 alters the internal state. Consistent with primacy effects (Asch, 1946[24]) and attitude polarization (Lord, Ross & Lepper, 1979[4]).

> **H3 (Interval Effect):** Repeated information should produce different effects depending on the temporal spacing between events. Consistent with findings on memory reconsolidation (Nader, Schafe & LeDoux, 2000[17]).

> **H4 (Differential Persistence):** Beliefs and attitudes should show greater temporal persistence than individual thoughts. Supported by the literature on hierarchical Bayesian learning rates (Behrens et al., 2007[10]; Lee & Mumford, 2003[21]).

> **H5 (Feedback Superiority):** A model with feedback from conscious experience should better explain temporal changes in subjective reports than a model that treats each information event independently. Consistent with Active Inference (Friston et al., 2010[26]; Friston, 2017[27]).

5. Possibilities of the MHST Model

5.1 Formal Description of State-Dependent Perception

The model provides a mathematical language to describe why identical information can produce different conscious representations. If two people with different cognitive states X_1 and X_2 receive the same information I, then CR_1 != CR_2, because the transformation function F produces different outputs. This prediction corresponds directly to the classic finding of the anchoring effect: Tversky and Kahneman (1974[1]) showed that judgments under uncertainty are systematically influenced by a reference point (anchor), whose adjustment is typically insufficient. In the MHST framework, the anchor can be interpreted as a component of B(t) or A(t) that modulates the processing of subsequent information. Likewise, confirmation bias — the tendency to seek, interpret, and recall information in a way that supports existing beliefs (Wason, 1960[2]; Nickerson, 1998[3]) — is a direct consequence of state-dependent processing: when B(t) is strongly pronounced, contradictory information is processed with lower weight.

5.2 Mathematical Formulation of Path Dependence

The model provides a formal proof of sequence sensitivity: reversing the presentation order of information yields a fundamentally different conscious outcome. This non-commutative property (CR_2' != CR_2) is a direct consequence of nonlinear transformation and state dependence. Lord, Ross, and Lepper (1979[4]) empirically demonstrated that participants with strong prior attitudes toward capital punishment interpreted the same mixed evidence in a way that reinforced their respective positions — a phenomenon they termed *biased assimilation* and *attitude polarization*. In the MHST framework, this corresponds to the prediction that different initial states X_0 lead to divergent trajectories in the cognitive state space given identical information input.

5.3 Multiscale Dynamics and Empirical Anchor Points

The model postulates different temporal dynamics for the five variables. This prediction is qualitatively supported by several empirical findings:

- **Sensory level (T analogue):** Mismatch negativity (MMN), an ERP marker for automatic prediction-error detection, peaks at approximately 150–250 ms post-stimulus (Wacongne et al., 2011[11]; Bendixen et al., 2012[12]). Wacongne et al. (2011) developed a neural model that explains the MMN as a consequence of Predictive Coding.
- **Precision modulation (SC analogue):** Feldman and Friston (2010[13]) formalized attention as precision weighting of prediction errors and showed that increasing precision by a factor of approximately 2–4x is sufficient to reproduce standard attentional cueing effects.
- **Belief updating (B analogue):** Behrens et al. (2007[10]) estimated Bayesian learning rates in a volatile probabilistic reversal task and reported values of approximately 0.3–0.7 (on a 0–1 scale), higher during volatile phases and lower during stable phases.
- **Memory reconsolidation (SC analogue):** Nader, Schafe, and LeDoux (2000[17]) showed that consolidated fear memories return to a labile state after retrieval and can be disrupted by protein synthesis inhibition. This supports the notion that conscious experience (CR) acts back on unconscious structures (SC) — loop 5 of the MHST model.

These values are *illustrative anchor points* from foreign paradigms, not parameter estimates for the MHST model itself. They merely show that the postulated multiscale structure is *consistent* with empirical data, not that it is *proven*.

5.4 Bridge to Established Theory

Mapping MHST variables onto the hierarchical Active Inference framework is scientifically productive, as it connects the model to an established, peer-reviewed theory. Friston (2010[5]) argued that the brain acts as a system that minimizes free energy. Clark (2013[6]) extended this into a comprehensive theory of cognitive science. Hohwy (2013[9]) developed the philosophical implications. Seth (2021[15]) generalized the framework to interoceptive perception and emotional experience. Through this mapping, the abstract functions F and G acquire a concrete mathematical form that can in principle be simulated computationally.

5.5 Clinical Application Framework

The model offers a formal framework for Computational Psychiatry (Montague et al., 2012[28]; Adams et al., 2013[29]). Sterzer et al. (2018[7]) argued that psychotic disorders can be understood as disorders of precision-weighted prediction-error processing. Friston and Stephan (2007[30]) proposed a dysconnection hypothesis of schizophrenia. The MHST model can formalize these disorders:

- **Schizophrenia:** Pathologically high prior precision (Pi_2 >> Pi_1) causes top-down predictions to override sensory input, producing hallucinations and delusions (Sterzer et al., 2018[7]; Friston & Stephan, 2007[30]).
- **Depression:** Rigid, low-volatility prior belief states with suppressed learning rates (kappa -> 0) prevent positive experiences from updating core beliefs. Seth and Friston (2016[14]) link this to interoceptive inference.
- **PTSD:** Trauma-induced precision spikes create deep attractor wells in the state space, which are reactivated by minimal cue stimuli. Consistent with the literature on memory reconsolidation (Nader et al., 2000[17]).

5.6 Computational Simulability

The discretized Euler equation of the MHST-PP model enables numerical simulations. The simulation contained in the synthesis document demonstrates how a two-level hierarchical network with an initial prior mu(0) = 3.0 and sensory input s = 10.0 converges to a posterior state via gradient descent, with free energy decreasing from 19.6 to 4.9 over 10 steps. The sensitivity analysis shows that under pathologically high prior precision, the state update remains minimal — a formal expression of *dogmatic* behavior. Such simulations are illustrative but do not constitute proof of the model's empirical validity.

6. Limitations of the MHST Model

6.1 Lack of Construct Validity

The five variables T, B, A, SC, and CR are theoretical constructs without established operationalization. In the sense of construct validity (Cronbach & Meehl, 1955[16]), a nomological network of theoretical relationships supported by empirical data is missing. No study has measured T(t), B(t), A(t), SC(t), or CR(t) as they are defined in the MHST model. In particular, SC (unconscious processes) is by definition not introspectively accessible and can only be inferred indirectly — through the process-dissociation procedure (Jacoby, 1991[8]) or subliminal priming paradigms. Measuring CR(t) through introspective self-reports is itself reactive: the act of verbalization can alter the reported thought (Schwarz, 1999[19]).

6.2 Parameters from Foreign Paradigms

The numerical values in the model are not parameter estimates for the MHST model. Three reasons argue against direct use: (1) **Different measured quantities** — MMN latency measures neural reaction times, not the value of T(t); Behrens et al.'s learning rate measures trial-by-trial behavioral updating, not B(t). (2) **Paradigm specificity** — all values are specific to their experimental paradigms. (3) **No joint estimation** — even within PP research, parameters are typically estimated jointly, per study, per hierarchical level.

6.3 Incompleteness of the Five-Variable Architecture

The five-variable architecture is necessarily incomplete. It neglects important additional variables:

- **Memory** — The model contains no explicit memory module, even though memories substantially influence the interpretation of new information. The reconsolidation literature (Nader et al., 2000[17]) shows that retrieved memories are modifiable.
- **Emotion** — Affective states are not fully captured in T or A. Seth and Friston (2016[14]) argue that emotions should be understood as interoceptive Active Inference, which requires an additional processing level.
- **Attention** — Although SC is partly interpreted as precision weighting, attention as an independent resource is not explicitly modeled (Feldman & Friston, 2010[13]).
- **Physiological state** — Hunger, fatigue, and hormonal balance influence cognitive processing but are not included (Seth, 2013[31]; Seth & Friston, 2016[14]).
- **Social context** — Social interactions, cultural background, and group dynamics are entirely absent.
- **Phenomenal consciousness** — The model identifies CR(t) with the converged posterior, but offers no explanation for why certain states are accompanied by phenomenal consciousness (the *hard problem of consciousness*, Chalmers, 1995[32]).

6.4 Limited Predictive Value

In its current form, the model is descriptive, not predictive. Without estimated parameters, it cannot make quantitative predictions about specific perceptual events. The linear approximation (Equation 22) contains unestimated weights w_i, and the model itself warns that a realistic system would require nonlinear interaction terms. The claim that the model represents a 'complete, mathematically unified theory of human cognition' is scientifically overstated.

6.5 Criticism of the Free-Energy Principle

The underlying FEP is not without controversy. Critics point out that the principle is partly *unfalsifiable*: if every biological system minimizes free energy by definition, no observation can refute the principle. Colombo (2014[33]) argues that the FEP is useful as a heuristic framework but does not qualify as an empirical theory in the strict sense. This criticism applies equally to the MHST model.

7. Explanation of the Scientific Sources

This work draws on 35 scientific sources, which are explained below. They can be grouped into six thematic categories:

7.1 Heuristics and Cognitive Biases

[1] **Tversky & Kahneman (1974)** — *Judgment under uncertainty: Heuristics and biases*. Science, 185(4157), 1124–1131.

This article is one of the most influential in the history of psychology. Tversky and Kahneman showed that people systematically use heuristics under uncertainty — mental shortcuts that often lead to errors. Three heuristics were identified: representativeness (judgment by similarity), availability (judgment by ease of recall), and anchoring-and-adjustment (judgment from a reference point). The anchoring effect is directly relevant to the MHST model: it shows that the prior cognitive state (B(t) or A(t)) systematically biases the processing of new information.

[2] **Wason (1960)** — *On the failure to eliminate hypotheses in a conceptual task*. Quarterly Journal of Experimental Psychology, 12(3), 129–140.

Wason's classic 2-4-6 experiment demonstrated confirmation bias: participants tend to seek hypotheses that confirm their guess rather than testing hypotheses that could refute it. In the MHST model, this corresponds to the prediction that a strong belief state B(t) reduces the processing of contradictory information.

[3] **Nickerson (1998)** — *Confirmation bias: A ubiquitous phenomenon in many guises*. Review of General Psychology, 2(2), 175–220.

Nickerson's comprehensive review article summarizes four decades of research on confirmation bias. He argues that confirmation bias is not a sign of insufficient intelligence but a fundamental feature of human thinking. This perspective supports the MHST prediction of state dependence (H1).

[4] **Lord, Ross & Lepper (1979)** — *Biased assimilation and attitude polarization*. J. Pers. Soc. Psychol., 37(11), 2098–2109.

This study showed that participants with strong attitudes toward capital punishment interpreted the same mixed evidence in a way that reinforced their respective positions. This is direct evidence for the MHST prediction of path dependence (H2): different initial states lead to divergent trajectories given identical information.

7.2 Predictive Processing and Free Energy

[5] **Friston (2010)** — *The free-energy principle: a unified brain theory?* Nature Reviews Neuroscience, 11(2), 127–138.

This is the foundational paper of the Free-Energy Principle. Friston argues that biological systems minimize free energy — an information-theoretic measure that provides an upper bound on surprise. The principle unifies perception, action, and learning under a single mathematical framework. For the MHST model, this is the theoretical foundation: CR(t) is defined as the argmin of free energy, and G is operationalized as gradient descent.

[6] **Clark (2013)** — *Whatever next? Predictive brains, situated agents, and the future of cognitive science*. Behavioral and Brain Sciences, 36(3), 181–204.

Clark extends the FEP into a comprehensive theory of cognitive science. He argues that the brain is essentially a prediction machine that constantly generates models of the world and adapts them to sensory data. This article provides the conceptual bridge between the abstract MHST model and the concrete mechanics of Predictive Processing.

[9] **Hohwy (2013)** — *The Predictive Mind*. Oxford University Press.

Hohwy's book develops the philosophical implications of Predictive Processing. He argues that consciousness should be understood as an interpretive process in which the brain generates and tests hypotheses about the causes of sensory data. For the MHST model, this provides the philosophical foundation for defining CR(t) as conscious representation.

[15] **Seth (2021)** — *Being You: A New Science of Consciousness*. Faber & Faber.

Seth generalizes the Predictive Processing framework to interoceptive perception — the perception of internal bodily states. He argues that emotions should be understood as predictions about bodily state. This extends the MHST model with a dimension missing from the original five-variable architecture.

[26] **Friston et al. (2010)** — *Action and behavior: A free-energy formulation*. Biological Cybernetics, 102(3), 227–260.

This paper extends the FEP to motor control and behavior. Active Inference means that an agent not only adjusts its internal states but also actively changes its environment to minimize prediction errors. In the MHST model, this corresponds to the five feedback loops through which conscious experience acts back on the cognitive state.

[27] **Friston (2017)** — *Active inference and the free-energy principle*. In: Computational Psychiatry, 129–152.

Friston summarizes the Active Inference framework and extends it with formal models of action selection. This work provides the mathematical foundation for the feedback loops in the MHST model, in particular the consolidation of beliefs (Equation 18) and attitudes (Equation 19).

7.3 Hierarchical Processing and Neural Implementation

[10] **Behrens et al. (2007)** — *Learning the value of information in an uncertain world*. Nature Neuroscience, 10(9), 1214–1221.

Behrens et al. estimated Bayesian learning rates in a volatile probabilistic reversal task. They found learning rates of approximately 0.3–0.7, which adapted to the volatility of the environment. In the MHST model, these values serve as an illustrative anchor point for the time constant of B(t) (beliefs).

[11] **Wacongne, Changeux & Dehaene (2011)** — *A neuronal model of predictive coding accounting for the mismatch negativity*. Journal of Neuroscience, 31(49), 17838–17847.

Wacongne et al. developed a neural model that explains mismatch negativity (MMN) as a consequence of Predictive Coding. The MMN is an ERP that occurs when a sensory stimulus does not match the prediction. This provides the neurophysiological anchor point for the time constant of T(t) (thoughts): 150–250 ms.

[12] **Bendixen, SanMiguel & Schroger (2012)** — *Early electrophysiological indicators for predictive processing in audition*. International Journal of Psychophysiology, 83(2), 120–131.

Bendixen et al. summarize the evidence for early electrophysiological markers of predictive processing in the auditory system. Their work supports the MHST prediction that prediction errors are detected on a fast timescale (milliseconds).

[13] **Feldman & Friston (2010)** — *Attention, uncertainty, and free-energy*. Frontiers in Human Neuroscience, 4, 215.

Feldman and Friston formalize attention as precision weighting of prediction errors. They show that increasing precision by a factor of 2–4x is sufficient to reproduce standard attentional cueing effects. In the MHST model, this is the operationalization of SC(t) as the precision matrix Pi.

[20] **Bastos et al. (2012)** — *Canonical microcircuits for predictive coding*. Neuron, 76(4), 695–711.

Bastos et al. identified the canonical microcircuits of the neocortex that implement Predictive Coding: prediction errors in superficial pyramidal cells (layers II/III) and predictions in deep pyramidal cells (layers V/VI). This provides the neuroanatomical foundation for the hierarchical structure of the MHST model.

[21] **Lee & Mumford (2003)** — *Hierarchical Bayesian inference in the visual cortex*. JOSA A, 20(7), 1434–1448.

Lee and Mumford proposed that the visual cortex functions as a hierarchical Bayesian inference system. Each cortical level generates predictions for the level below and updates itself based on prediction errors. This is the conceptual foundation for MHST equations 8–12.

[22] **Rao & Ballard (1999)** — *Predictive coding in the visual cortex*. Nature Neuroscience, 2(1), 79–87.

Rao and Ballard laid the foundation for Predictive Coding in the visual system. They showed that a hierarchical model sending predictions downward and prediction errors upward can reproduce many properties of the visual cortex, including endstopping and extra-classical receptive fields.

[23] **Yu & Dayan (2005)** — *Uncertainty, neuromodulation, and attention*. Neuron, 46(4), 681–692.

Yu and Dayan connect neuromodulatory systems (acetylcholine, noradrenaline) to precision control in Predictive Coding. Acetylcholine signals expected uncertainty (top-down), noradrenaline signals unexpected uncertainty (bottom-up). In the MHST model, this provides the biological basis for the modulation of SC(t) by attention.

7.4 Construct Validity, Measurement, and Memory

[16] **Cronbach & Meehl (1955)** — *Construct validity in psychological tests*. Psychological Bulletin, 52(4), 281–302.

This is the foundational article on construct validity. Cronbach and Meehl argue that a psychological construct must be validated through a nomological network — a web of theoretical relationships supported by empirical data. For the MHST model, this is the central criticism: its five variables are constructs without established validation.

[8] **Jacoby (1991)** — *A process dissociation framework*. Journal of Memory and Language, 30(5), 513–541.

Jacoby developed the process-dissociation procedure, a method for separating conscious and unconscious cognitive processes. This is relevant to the MHST model, since SC(t) (unconscious processes) is not introspectively accessible and can only be measured indirectly.

[17] **Nader, Schafe & LeDoux (2000)** — *The labile nature of consolidation theory*. Nature Reviews Neuroscience, 1(3), 216–219.

Nader et al. showed that consolidated fear memories return to a labile state after retrieval and can be disrupted by protein synthesis inhibition. This supports the MHST prediction that conscious experience (CR) acts back on unconscious structures (SC) — loop 5 of the model.

[19] **Schwarz (1999)** — *Self-reports: How the questions shape the answers*. American Psychologist, 54(2), 93–105.

Schwarz shows that introspective self-reports are reactive: the way questions are asked influences the answers. For the MHST model, this is a methodological challenge, since CR(t) is typically measured through self-reports, which can themselves alter the conscious state.

[34] **Greenwald, McGhee & Schwartz (1998)** — *Measuring individual differences in implicit cognition: The IAT*. Journal of Personality and Social Psychology, 74(6), 1464–1480.

Greenwald et al. developed the Implicit Association Test (IAT), a measure of implicit attitudes that does not rely on introspective self-reports. The IAT could serve as a measurement instrument for A(t) in the MHST model.

[35] **Csikszentmihalyi & Larson (1987)** — *Validity and reliability of the Experience-Sampling Method*. Journal of Nervous and Mental Disease, 175(9), 526–536.

Csikszentmihalyi and Larson developed the Experience-Sampling Method (ESM), in which participants are randomly surveyed multiple times daily about their current thoughts and feelings. The ESM could serve as a measurement instrument for T(t) and CR(t) in the MHST model.

7.5 Clinical and Philosophical Perspectives

[7] **Sterzer et al. (2018)** — *The predictive coding account of psychosis*. Biological Psychiatry, 84(9), 634–643.

Sterzer et al. argue that psychotic disorders can be understood as disorders of precision-weighted prediction-error processing. In schizophrenia, pathologically high prior precision causes top-down predictions to override sensory input, producing hallucinations. In the MHST model, this can be formalized as $Pi_2 >> Pi_1$.

[14] **Seth & Friston (2016)** — *Active interoceptive inference and the emotional brain*. Philosophical Transactions of the Royal Society B, 371(1708), 20160007.

Seth and Friston extend Active Inference to interoceptive perception — the perception of internal bodily states. They argue that emotions should be understood as predictions about bodily state. This extends the MHST model with an emotional dimension missing from the original architecture.

[28] **Montague et al. (2012)** — *Computational psychiatry*. Trends in Cognitive Sciences, 16(1), 72–80.

Montague et al. found the field of Computational Psychiatry, which understands psychiatric disorders through computational models. The MHST model can serve as a framework for this approach, in particular through its ability to formally describe clinical conditions.

[29] **Adams, Shipp & Friston (2013)** — *Predictions not commands: Active inference in the motor system*. Brain Structure and Function, 218(3), 611–643.

Adams et al. apply Active Inference to the motor system. They argue that motor control should be understood not as command execution but as the minimization of prediction errors through muscle activation. This extends the scope of the MHST model to motor behavior.

[30] **Friston & Stephan (2007)** — *Free-energy and the brain*. Synthese, 159(3), 417–458.

Friston and Stephan formulate the dysconnection hypothesis of schizophrenia: the disorder is based not on local defects but on disrupted connectivity between hierarchical levels. In the MHST model, this corresponds to disrupted precision weighting Pi_l(t) between levels.

[31] **Seth (2013)** — *Interoceptive inference, emotion, and the embodied self*. Trends in Cognitive Sciences, 17(11), 565–573.

Seth develops a theory of the sense of self based on interoceptive inference. The sense of self thus arises as a prediction about one's own bodily state. For the MHST model, this provides an extension with physiological states missing from the original architecture.

[32] **Chalmers (1995)** — *Facing up to the problem of consciousness*. Journal of Consciousness Studies, 2(3), 200–219.

Chalmers formulates the 'hard problem of consciousness': why does subjective experience exist at all? The MHST model identifies CR(t) with the converged posterior but does not explain why this state is accompanied by phenomenal consciousness. Chalmers' work marks this limit.

[33] **Colombo (2014)** — *Deep and beautiful? The reward prediction error hypothesis of dopamine*. Studies in History and Philosophy of Science Part C, 45, 57–67.

Colombo criticizes the Free-Energy Principle as partly unfalsifiable: if every biological system minimizes free energy by definition, no observation can refute the principle. This criticism also applies to the MHST model, which builds on the FEP.

7.6 Attitude and Judgment Formation

[24] **Asch (1946)** — *Forming impressions of personality*. Journal of Abnormal and Social Psychology, 41(3), 258–290.

Asch demonstrated primacy effects in impression formation: information presented first has a stronger influence on the overall impression than later information. This is direct evidence for the MHST prediction of path dependence (H2): the order of information influences the outcome.

[25] **Hogarth & Einhorn (1992)** — *Order effects in belief updating*. Cognitive Psychology, 24(1), 1–55.

Hogarth and Einhorn systematically investigated how the order of information influences judgment formation. They found that primacy effects dominate with short information sequences, while recency effects dominate with long sequences. This refines the MHST prediction of cognitive hysteresis.

8. Discussion

The MHST model occupies a traditional tension in cognitive science between conceptual depth and empirical testability. Its strength lies in its ability to integrate phenomena that have been known in psychology for decades — anchoring effects (Tversky & Kahneman, 1974[1]), confirmation bias (Wason, 1960[2]; Nickerson, 1998[3]), attitude polarization (Lord, Ross & Lepper, 1979[4]), primacy effects (Asch, 1946[24]) — into a unified mathematical framework. The formalization through Predictive Processing connects it to one of the most influential theories in contemporary cognitive science (Friston, 2010[5]; Clark, 2013[6]).

The central weakness is a gap in construct validity: the model describes relationships between constructs that have not themselves been measured. This is not a unique defect of the MHST model — many influential models in cognitive science share this feature — but it limits the current scientific value to the conceptual and theoretical domain.

A realistic first step toward validation would not be the simultaneous operationalization of all five variables, but restricting attention to two or three variables that are most accessible to established measurement methods. For B(t), Bayesian updating paradigms, such as those developed by Behrens et al. (2007[10]), offer a path forward. For A(t), decades-old psychometrically tested scales as well as implicit measures such as the Implicit Association Test (Greenwald et al., 1998[34]) are available. For CR(t), psychophysical judgments and confidence ratings are suitable. A controlled two-event experiment could test whether CR_2 actually depends on I_1 (Hypothesis H2).

Critically, the claim that the model represents a 'complete, mathematically unified theory of human cognition' is scientifically overstated and is refuted by the model's own stated limitations, which explicitly warn against presenting illustrative literature values or proposed procedures as a validated, operationalized, and empirically tested version of the framework.

9. Conclusion

The MHST model is an intellectually ambitious and mathematically specifiable conceptual framework for understanding subjective perception as a temporally dynamic, state-dependent, and path-dependent process. Its theoretical coherence is secured through its connection to the established Predictive Processing framework (Friston, 2010[5]; Clark, 2013[6]; Hohwy, 2013[9]). Its central predictions — state dependence, path dependence, and multiscale dynamics — are consistent with a broad empirical literature ranging from classic findings in cognitive psychology (Tversky & Kahneman, 1974[1]; Wason, 1960[2]) to neurophysiological markers (Wacongne et al., 2011[11]; Behrens et al., 2007[10]).

Its limitations lie in the lack of construct validity of its variables (Cronbach & Meehl, 1955[16]), the absence of estimated parameters, the incompleteness of the five-variable architecture, and the lack of empirical validation. The model is currently a theoretical architectural pattern that must be operationalized, mathematically refined, and empirically tested. Its value lies in its role as a conceptual foundation for future research in cognitive science, psychology, computational modeling, and clinical psychiatry — not as a finished theory.

10. Literaturverzeichnis

35 zitierte Quellen (geordnet nach Erstnennung)

[1] Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131. Link

[2] Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129-140.

[3] Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220.

[4] Lord, C. G., Ross, L. & Lepper, M. R. (1979). Biased assimilation and attitude polarization. *J. Pers. Soc. Psychol.*, 37(11), 2098-2109.

[5] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138. Link

[6] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.*, 36(3), 181-204.

[7] Sterzer, P. et al. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, 84(9), 634-643. Link

[8] Jacoby, L. L. (1991). A process dissociation framework. *J. Mem. Lang.*, 30(5), 513-541.

[9] Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

[10] Behrens, T. E. J. et al. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9), 1214-1221.

[11] Wacongne, C., Changeux, J.-P. & Dehaene, S. (2011). A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.*, 31(49), 17838-17847.

[12] Bendixen, A., SanMiguel, I. & Schroger, E. (2012). Early electrophysiological indicators for predictive processing in audition. *Int. J. Psychophysiol.*, 83(2), 120-131.

[13] Feldman, H. & Friston, K. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.*, 4, 215.

[14] Seth, A. K. & Friston, K. (2016). Active interoceptive inference and the emotional brain. *Phil. Trans. R. Soc. B*, 371(1708), 20160007.

[15] Seth, A. (2021). *Being You: A New Science of Consciousness*. Faber & Faber.

[16] Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.*, 52(4), 281-302. Link

[17] Nader, K., Schafe, G. E. & LeDoux, J. E. (2000). The labile nature of consolidation theory. *Nat. Rev. Neurosci.*, 1(3), 216-219. Link

[18] Jacoby, L. L. (1991). A process dissociation framework. *J. Mem. Lang.*, 30(5), 513-541.

[19] Schwarz, N. (1999). Self-reports: How the questions shape the answers. *Am. Psychol.*, 54(2), 93-105.

[20] Bastos, A. M. et al. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695-711.

[21] Lee, T. S. & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *JOSA A*, 20(7), 1434-1448.

[22] Rao, R. P. N. & Ballard, D. H. (1999). Predictive coding in the visual cortex. *Nat. Neurosci.*, 2(1), 79-87.

[23] Yu, A. J. & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4), 681-692.

[24] Asch, S. E. (1946). Forming impressions of personality. *J. Abnorm. Soc. Psychol.*, 41(3), 258-290.

[25] Hogarth, R. M. & Einhorn, H. J. (1992). Order effects in belief updating. *Cogn. Psychol.*, 24(1), 1-55.

[26] Friston, K. J. et al. (2010). Action and behavior: A free-energy formulation. *Biol. Cybern.*, 102(3), 227-260.

[27] Friston, K. J. (2017). Active inference and the free-energy principle. In: *Computational Psychiatry*, 129-152.

[28] Montague, P. R. et al. (2012). Computational psychiatry. *Trends Cogn. Sci.*, 16(1), 72-80.

[29] Adams, R. A., Shipp, S. & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Struct. Funct.*, 218(3), 611-643.

[30] Friston, K. J. & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3), 417-458.

[31] Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.*, 17(11), 565-573.

[32] Chalmers, D. J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.*, 2(3), 200-219.

[33] Colombo, M. (2014). Deep and beautiful? The reward prediction error hypothesis of dopamine. *Stud. Hist. Philos. Sci. C*, 45, 57-67.

[34] Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The IAT. *J. Pers. Soc. Psychol.*, 74(6), 1464-1480.

[35] Csikszentmihalyi, M. & Larson, R. (1987). Validity and reliability of the Experience-Sampling Method. *J. Nerv. Ment. Dis.*, 175(9), 526-536.