# Behavioral & Neurobiological Modeling Parameters: Data Synthesis

This document outlines six core domains of human behavior and neurobiology. It details the source literature, key quantitative evidence, and specific variables required to model these states effectively. Each section is paired with relevant datasets for computational modeling.

## 1. Threat, Emotion & Physiology

**Primary Focus:** The biological construction of emotion and the autonomic response to fear.

### Core Mechanism & Evidence

* **Prediction Error:** According to the Theory of Constructed Emotion, the brain predicts threats rather than passively detecting them.
* **Skin Conductance (SCL):** A reliable predictor of arousal; 95% of subjects demonstrate elevated SCL prior to a threat event.
* **Heart Rate (HR):** Contrary to "fight or flight," 80% (28/35 subjects) experience bradycardia (heart rate slowing) during threat anticipation.
* **Visual Scanning:** Trait anxiety correlates with "hyperscanning" (rapid, excessive eye movement) to identify danger.

### Modeling Variables

* **Input Vectors:** Visual Stimuli (Height/Predator) + Interoception (Body state).
* **Internal State:** Coordinates of Valence (Pleasant/Unpleasant) and Arousal (High/Low).

### Relevant Datasets

* [**VR Emotion Multimodal Dataset**](https://www.kaggle.com/datasets/programmer3/vr-emotion-multimodal-dataset)
  + *Description:* Multimodal VR emotion dataset containing EEG, GSR/EDA, PPG/heart rate, eye‑tracking features, and self‑reported valence/arousal under different emotional states.
* [**WESAD (Wearable Stress and Affect Detection)**](https://www.kaggle.com/datasets/orvile/wesad-wearable-stress-affect-detection-dataset)
  + *Description:* Multimodal wearable dataset with ECG (HR), EDA, respiration, temperature, accelerometer, etc., labeled as neutral, stress, and amusement. (See also: [Full Dataset Code](https://www.kaggle.com/datasets/mohamedasem318/wesad-full-dataset/code)).
* [**EEG Brainwave Dataset: Feeling Emotions**](https://www.kaggle.com/datasets/birdy654/eeg-brainwave-dataset-feeling-emotions)
  + *Description:* EEG recordings labeled with positive vs negative emotions, suitable for mapping brain activity to valence.
* [**Emotions and Heart Rate Scale Classification**](https://www.kaggle.com/datasets/isameeramohamed/emotions-and-heart-rate-scale-classification)
  + *Description:* Emotion dataset with heart‑rate and EEG‑derived features and valence/arousal labels, aligned with DEAP‑style annotation.
* [**Multimodal Emotion Recognition Dataset**](https://www.kaggle.com/datasets/ziya07/multimodal-emotion-recognition-dataset)
  + *Description:* Emotion recognition using multimodal signals (physiological + behavioral) collected as people interact with digital artworks.

## 2. Decision Making & Cognitive Bias

**Primary Focus:** The architecture of judgment and system latency.

### Core Mechanism & Evidence

* **Dual Process Theory:** System 1 is automatic/heuristic; System 2 is effortful/logical.
* **Framing Effects:** Decisions are biased based on "Loss" vs. "Gain" presentation, even with identical outcomes.
* **Substitution:** The brain substitutes computationally difficult questions (e.g., life satisfaction) with easier ones (e.g., current mood).

### Modeling Variables

* **Processing Mode:** System 1 (Intuitive) vs. System 2 (Rational).
* **Accessibility Heuristic:** Choice probability is weighted by the ease of recall rather than statistical likelihood.

### Relevant Datasets

* [**Cognitive Distortion Detection Dataset**](https://www.kaggle.com/datasets/sagarikashreevastava/cognitive-distortion-detetction-dataset)
  + *Description:* Patient–therapist Q&A annotated with cognitive distortions, ideal for modeling biased reasoning and substitution heuristics.
* [**NLP Mental Health Conversations**](https://www.kaggle.com/datasets/thedevastator/nlp-mental-health-conversations)
  + *Description:* Anonymized conversations between users and psychologists, enabling analysis of emotion‑driven, heuristic decision talk.
* [**Risk Behavior Features Analysis**](https://www.kaggle.com/datasets/doublevvvvv/risk-behavior-features-analysis)
  + *Description:* Questionnaire‑based dataset with demographics and financial status suitable for modeling risk preferences.
* [**Credit Risk Dataset**](https://www.kaggle.com/datasets/laotse/credit-risk-dataset)
  + *Description:* Classic credit‑default classification dataset reflecting institutional decision rules and thresholds.
* [**News Dataset for News Bias Analysis**](https://www.kaggle.com/datasets/articoder/news-dataset-for-news-bias-analysis)
  + *Description:* News articles curated to analyze media bias across outlets, relevant to framing effects.
* [**CrowS-Pairs (Social Biases in MLMs)**](https://www.kaggle.com/datasets/thedevastator/a-dataset-for-measuring-social-biases-in-mlms)
  + *Description:* Sentence pairs measuring social biases, reflecting how stereotypical associations shape judgments.
* [**Psychology Datasets Collection**](https://www.kaggle.com/datasets?tags=11211-Psychology)
  + *Description:* A curated entry point to 33+ psychology‑related datasets. (See also: [Discussion Thread](https://www.kaggle.com/general/304994)).

## 3. Flow, Motivation & Optimal Experience

**Primary Focus:** The psychology of engagement and the disconnect between satisfaction and desire.

### Core Mechanism & Evidence

* **Flow State:** Optimal experience occurs when Task Challenge (C) matches Agent Skill (S).
* **The Work Paradox:** 54% of flow states occur at work (high quality of experience), yet workers report lower motivation than in leisure (where only 18% of flow occurs).

### Modeling Variables

* **Flow Ratio:** High C + High S.
* **State Determination:**
  + Anxiety: C > S
  + Boredom: S > C
  + Flow: S \approx C

### Relevant Datasets

* [**HF 2021 Employee Engagement Dataset (Happyforce)**](https://www.kaggle.com/datasets/harriken/myhappyforce-survey-employee-stress)
  + *Description:* Time‑series survey data on employee engagement, stress, and sentiment, ideal for modeling dynamic flow‑like states.
* [**Employee Satisfaction Index Dataset**](https://www.kaggle.com/datasets/mohamedharris/employee-satisfaction-index-dataset)
  + *Description:* Detailed job satisfaction index across multiple aspects of work, aligning with quality of experience.
* [**Employee Performance and Productivity Data**](https://www.kaggle.com/datasets/mexwell/employee-performance-and-productivity-data/data)
  + *Description:* HR dataset with performance and productivity metrics, useful for mapping C vs S via workload and performance.
* [**Work Productivity (Garment Workers)**](https://www.kaggle.com/datasets/vimalkumarnarasiman/work-productivity)
  + *Description:* Daily production‑line data supporting flow vs. overload vs. underload modeling in factory settings.
* [**AI Developer Productivity Dataset**](https://www.kaggle.com/datasets/atharvasoundankar/ai-developer-productivity-dataset)
  + *Description:* Simulated data over 500 days with deep‑work hours, distractions, and fatigue, closely tied to flow processes.
* [**Time Management and Productivity Insights**](https://www.kaggle.com/datasets/hanaksoy/time-management-and-productivity-insights)
  + *Description:* Captures time allocation, focus, and procrastination for productivity optimization.

## 4. Habit Formation & Automaticity

**Primary Focus:** The temporal dynamics required to shift behavior from effortful to automatic.

### Core Mechanism & Evidence

* **Asymptotic Growth:** Habit formation follows a non-linear curve; rapid initial gain followed by a plateau.
* **Time to Habit:** The median is 66 days, with a range of 18 to 254 days.
* **Resilience:** Missing a single day has no significant impact on the long-term habit curve.

### Modeling Variables

* **Curve Shape:** Asymptotic/Non-linear.
* **Complexity:** Time to habit scales with task complexity.
* **Consistency:** Discrete failures do not reset accumulated habit strength.

### Relevant Datasets

* [**Daily Habit Tracker Dataset**](https://www.kaggle.com/datasets/prince7489/daily-habit-tracker-dataset)
  + *Description:* Daily records of sleep, activity, and mood, suitable for estimating habit strength trajectories.
* [**Personal Productivity Tracker**](https://www.kaggle.com/datasets/mertaydin30/personal-productivity-tracker/data)
  + *Description:* Weekly tracking of habits and productivity over 12 weeks, suitable for estimating asymptotic habit growth curves.
* [**Student Study Habits**](https://www.kaggle.com/datasets/prekshad2166/student-study-habits)
  + *Description:* 500 students with weekly study hours and assignments, supporting habit complexity modeling.
* [**StudySmart: Student Habits vs Performance**](https://www.kaggle.com/datasets/sumedh1507/student-study-habits-dataset)
  + *Description:* Captures daily study behaviors and routines along with academic performance.
* [**Disease Risk from Daily Habits**](https://www.kaggle.com/datasets/mahdimashayekhi/disease-risk-from-daily-habits)
  + *Description:* 100,000 individuals with lifestyle features, useful for modeling long‑term consequences of stable habits.
* [**Overstimulation Behavior and Lifestyle Dataset**](https://www.kaggle.com/datasets/miadul/overstimulation-behavior-and-lifestyle-dataset)
  + *Description:* Features like screen time and sensory load, supporting modeling of habitual overexposure.

## 5. Sleep, Memory & Plasticity

**Primary Focus:** The neurochemical requirements for learning and memory encoding.

### Core Mechanism & Evidence

* **Memory Consolidation:** Sleep is an active process transferring data from the Hippocampus (short-term) to the Neocortex (long-term).
* **Performance Gain:** Sleep alone yields a 20–30% improvement in motor skill speed/accuracy.
* **Deprivation Cost:** Lack of sleep reduces hippocampal activation, causing failure to encode new episodic memories.

### Modeling Variables

* **Agent State:** Sleep Deprived (<7 hrs) vs. Well-rested.
* **Process Flow:** Info Transfer (Hippocampus and Neocortex).

### Relevant Datasets

* [**Sleep Health and Lifestyle Dataset (uom)**](https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset)
  + *Description:* 373–400 individuals with sleep duration, quality, and stress levels. (See also: [Alternate Version](https://www.kaggle.com/datasets/henryshan/sleep-health-and-lifestyle/data)).
* [**Sleep Efficiency Dataset**](https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency)
  + *Description:* Dataset focused on sleep efficiency and stages, used for modeling how sleep architecture affects health.
* [**Student Sleep Patterns Dataset**](https://www.kaggle.com/datasets/arsalanjamal002/student-sleep-patterns/data)
  + *Description:* Students’ sleep duration and quality, supporting studies of sleep vs. functioning in young adults.
* [**Sleep Deprivation & Cognitive Performance**](https://www.kaggle.com/datasets/sacramentotechnology/sleep-deprivation-and-cognitive-performance)
  + *Description:* Directly links sleep hours and daytime sleepiness to cognitive performance metrics.
* [**EEG Sleep and Awake Datasets**](https://www.kaggle.com/datasets/aidenmerker/eeg-sleep-and-awake-datasets)
  + *Description:* EEG time‑series labeled as sleep vs awake, allowing state classification.
* [**Siena Sleep EEG Dataset**](https://www.kaggle.com/datasets/ucimachinelearning/siena-sleep-eeg-dataset)
  + *Description:* Sleep EEG dataset oriented to epilepsy detection and sleep staging.
* [**EEG-Based Real-Time Sleep Stage Classifier**](https://www.kaggle.com/datasets/jocelyndumlao/eeg-based-real-time-sleep-stage-classifier/data)
  + *Description:* Data and labels used for a single‑channel EEG real‑time sleep staging system.

## 6. The Social Brain: Hierarchy & Status

**Primary Focus:** The neural drivers of social interaction and the biological costs of rank.

### Core Mechanism & Evidence

* **Social Threat:** The brain processes social threats (exclusion, status drop) with the same intensity as physical pain.
* **Cognitive Cost:** Social threat reduces Prefrontal Cortex (PFC) capacity, impairing problem-solving.
* **Hierarchy Gradient:** Health generally improves linearly with rank, except for "Alpha" males in unstable hierarchies who exhibit high cortisol (stress).

### Modeling Variables

* **SCARF Domains:** Status, Certainty, Autonomy, Relatedness, Fairness.
* **Social State:** In-group (Reward) vs. Out-group (Threat).  
  Relevant Datasets
* [**Employees Stress Level Dataset**](https://www.kaggle.com/datasets/chanchalagorale/employees-stress-level-dataset)
  + *Description:* Synthetic data on software employees with job satisfaction and stress levels, allowing modeling of workplace social threat.
* [**Employee Productivity and Satisfaction HR Data**](https://www.kaggle.com/datasets/adityaab1407/employee-productivity-and-satisfaction-hr-data)
  + *Description:* HR dataset with position (status proxy), productivity scores, and feedback scores.
* [**Employee Satisfaction Survey Data**](https://www.kaggle.com/datasets/redpen12/employees-satisfaction-analysis)
  + *Description:* Detailed survey covering multiple job facets, mapping directly onto SCARF‑like dimensions.
* [**Employee/HR Dataset (All in One)**](https://www.kaggle.com/datasets/ravindrasinghrana/employeedataset)
  + *Description:* Aggregated HR records and sentiment, helpful for modeling status and inclusion vs. well‑being.
* [**Health Outcomes and Socioeconomic Factors**](https://www.kaggle.com/datasets/thedevastator/uncovering-trends-in-health-outcomes-and-socioec)
  + *Description:* Regional dataset supporting hierarchy gradients (rank vs health) at a macro scale.
* [**Life Expectancy & Socio-Economic (World Bank)**](https://www.kaggle.com/datasets/mjshri23/life-expectancy-and-socio-economic-world-bank)
  + *Description:* Combines life expectancy with socio‑economic indicators, aligning with rank‑health relationships.
* [**Country Health Trends Dataset**](https://www.kaggle.com/datasets/sahirmaharajj/country-health-trends-dataset)
  + *Description:* Global dataset of health and socio‑economic indicators for macro‑level hierarchy modeling.