

# Personality-Modulated Trust Dynamics for Adaptive Control in Lower-Limb Wearable Robotics

---

## Executive Summary

---

This research develops a psychophysically-grounded computational framework for modeling human trust dynamics in lower-limb prosthetic/exoskeleton systems, with the ultimate goal of enabling trust-aware adaptive control. Unlike existing approaches that treat trust as a black box to be maximized, this work explicitly models how individual differences (Big Five personality traits, age) and perceptual mechanisms shape trust formation and evolution during human-robot physical interaction.

**Current Status:** Foundation infrastructure complete with validated perception models and personality integration. Next phase focuses on empirical validation and POMDP-based controller implementation.

---

## 1. Research Motivation

---

### 1.1 The Trust Modeling Gap

Current wearable robotics research acknowledges that trust is critical for effective human-robot collaboration, but most work:

- Treats trust as a scalar to be "maximized" without modeling underlying mechanisms
- Focuses on objective state mismatches (expected vs. actual performance) rather than subjective perceived states
- Ignores how individual differences and environmental context modulate trust dynamics
- Lacks integration between perceptual psychophysics and trust formation

### 1.2 Core Insight

**Trust is not directly observable by the robot and is fundamentally driven by perceived states, not objective states.** An action with 10% objective error might be perceived as:

- Catastrophic by a highly neurotic, elderly user in a high-risk context
- Acceptable by an agreeable, young user in a low-stakes situation

This variability must be explicitly modeled for effective personalized robot control.

### 1.3 Research Questions

1. How do individual differences (personality, age) quantitatively modulate perceived error magnitude in physical human-robot interaction?
  2. How does perceived error (not objective error) drive trust dynamics in a personality-dependent manner?
  3. Can a POMDP framework effectively estimate latent psychological state (trust, personality traits) from observable behavioral and physiological signals to enable adaptive control?
- 

## 2. Theoretical Framework

### 2.1 Three-Layer Trust Architecture

Trust is decomposed into three components following established HRI frameworks:

plaintext



$$T(t) = T_{\text{dispositional}} + T_{\text{situational}}(t) + T_{\text{learned}}(t)$$

#### T<sub>dispositional</sub> (Personality-Based Baseline)

- Static component derived from Big Five personality traits
- Agreeableness: Strongest positive predictor ( $r \approx 0.35$ )
- Neuroticism: Negative predictor ( $r \approx -0.30$ )
- Openness: Technology acceptance
- Conscientiousness: Reliability focus
- Extraversion: Social trust

#### T<sub>situational</sub> (Context-Dependent Modulation)

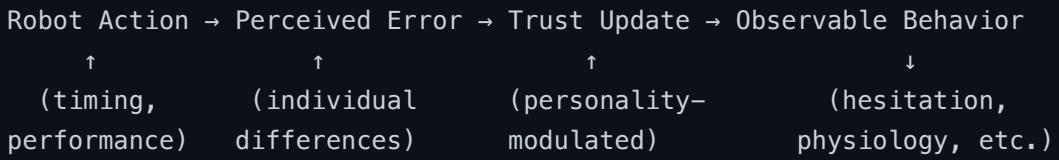
- Task risk/stakes (e.g., balance-critical gait phases)
- Environmental factors (noise, obstacles, terrain)
- User state (fatigue, cognitive load, anxiety)

#### T<sub>learned</sub> (Experience-Based Dynamics)

- Updated based on accumulated interaction experience
- Asymmetric: fast loss, slow recovery
- Personality-modulated learning rates

### 2.2 Perception-to-Trust Pipeline

plaintext



**Key Innovation:** Explicit modeling of the perception layer that mediates between objective robot performance and subjective trust formation.

### 3. Current Implementation: Infrastructure Components

#### 3.1 Psychophysically-Grounded Perception Model ✓

**Perceived Error Function:** Models how humans subjectively experience action errors based on:

plaintext



$$E = g \cdot \log(1 + |A|/A_{ref}) \cdot K(dt, \text{context}, \beta, t_0)$$

Where:

- **Magnitude term:** Weber-scaled force/torque deviation
- **Temporal kernel K:** Context-dependent (event vs. rhythmic tasks)
- **$\beta$  (sharpness):** Individual perceptual acuity → derived from Big Five
- **$t_0$  (temporal tolerance):** Age-dependent temporal binding window

**Theoretical Grounding:**

- Weber-Fechner law for magnitude scaling
- Temporal binding window from cognitive neuroscience
- Context-dependence for gait (event-driven) vs. rhythmic assistance

**Implementation Status:** Fully implemented and validated against psychophysical literature.

#### 3.2 Individual Differences Integration ✓

**Personality → Perceptual Parameters**

Implemented mapping from Big Five traits to perceptual sharpness ( $\beta$ ):

- **Openness** ( $w = 0.48$ ): Primary driver via cognitive ability pathway
- **Neuroticism** ( $w = -0.30$ ): Negative association with temporal discrimination

- **Conscientiousness, Extraversion, Agreeableness:** Smaller contributions

Evidence-weighted approach accounts for varying confidence levels in literature (moderate to insufficient direct evidence for temporal discrimination).

### Age → Temporal Tolerance

Elderly users have wider Temporal Binding Windows (TBW):

- Young adults (25y):  $t_0 \approx 30\text{ms}$
- Older adults (85y):  $t_0 \approx 100\text{ms}$
- Linear interpolation with age normalization

**Implementation Status:** Both mappings fully implemented with confidence weighting and age-based adjustments.

## 3.3 Trait-Observation Correlation Model

**Likelihood Matrix:** 11 observations × 5 personality traits

Observable behaviors/physiological signals correlated with personality:

- **Behavioral:** Hesitation, task refusal, weight transfer delay, monitoring frequency
- **Physiological:** Heart rate, palm sweat, muscle co-contraction

Matrix encodes  $P(\text{observation} \mid \text{personality, trust})$  based on HRI literature.

**Trust-Dependent Modulation:** Observations are modulated by current trust level using trust sensitivity weights (e.g., high trust → reduced monitoring).

**Implementation Status:** Correlation matrix defined, trust-modulation tensor computed, visualization tools implemented.

## 3.4 Biomechanically-Grounded Action Representation

**Robot Action Model** with two error dimensions:

### 1. Performance Error (functional quality):

- Walking symmetry (Symmetry Index: 0-100%)
- Balance stability ( $\phi$ -bonacci index: 0-100)
- Metabolic cost reduction (0-100%)
- Weighted combination → scalar performance error [0,1]

### 2. Timing Error (predictability):

- Temporal offset normalized to [0,1]
- Affects unpredictability independently of performance quality

**Combined Action Error:** Accounts for interaction between performance and timing failures.

**Implementation Status:** Fully implemented with Pydantic validation, random generation utilities, and manual override capabilities.

---

## 4. Technical Evaluation of Current Infrastructure

### 4.1 Strengths

#### 1. Psychophysically Rigorous

- Perception model grounded in Weber-Fechner law and temporal psychophysics
- Not ad-hoc: parameters justified by cognitive neuroscience literature
- Context-aware (event vs. rhythmic tasks)

#### 2. Individual Differences Are Concrete

- Big Five → perceptual sharpness mapping with evidence-weighted contributions
- Age → temporal tolerance based on empirical TBW data
- Not just "personality matters" hand-waving

#### 3. Observation Model Enables Inference

- Trait-observation correlation matrix provides  $P(\text{obs}|\text{trait}, \text{trust})$
- Foundation for Bayesian inference in POMDP (robot estimates hidden psych-state from observations)
- Visualization tools for model validation

#### 4. Biomechanically Relevant

- Action representation uses actual metrics from exoskeleton literature
- Walking symmetry, balance, metabolic cost are clinically meaningful
- Appropriate for lower-limb prosthesis/exoskeleton domain

#### 5. Modular and Extensible

- Clean separation: Human model, Action model, (future: Environment, Controller)
- Easy to add new observations, personality factors, or perception mechanisms
- Pydantic validation prevents invalid states during experimentation

### 4.2 Current Scope

This implementation represents a **sophisticated forward model** of human trust perception —ONE essential component of a complete POMDP-based controller, not the controller itself.

**What exists:** The "human simulator" that can generate trust trajectories and observations given robot actions.

**What this enables:**

- Hypothesis testing (e.g., "Do high-N individuals lose trust faster?")
- Synthetic data generation for controller development
- Parameter sensitivity analysis
- Foundation for empirical validation studies

### 4.3 Computational Considerations

- **State space:** Continuous (trust, personality traits, physical state)
  - **Observation space:** Mixed discrete (behavioral) + continuous (physiological)
  - **Tractability:** Current implementation is forward simulation—efficient for validation
  - **POMDP scaling:** Will require approximate inference (particle filters, variational methods)
- 

## 5. Future Research Steps

### Phase 1: Trust Dynamics Implementation (Weeks 1-4)

**Objective:** Complete the trust update mechanism to close the perception-trust loop.

**Tasks:**

1. Implement three-layer trust model (dispositional + situational + learned)
2. Define asymmetric learning rates ( $\alpha_{\text{loss}}$ ,  $\alpha_{\text{gain}}$ ) with personality modulation
3. Add memory/history tracking for cumulative effects
4. Integrate user state variables (fatigue, cognitive load)
5. Implement observation generation: sample observations based on current trust/traits

**Deliverables:**

- Functional human-robot interaction simulator
- Trust trajectory generation for different personalities  $\times$  robot behaviors
- Visualization tools for trust dynamics

### Phase 2: Model Validation via Simulation (Weeks 5-8)

**Objective:** Demonstrate that model produces theoretically coherent predictions.

**Experiments:**

1. **Personality effects:** Show high-N loses trust faster, high-A recovers faster

2. **Age effects:** Elderly users more tolerant of timing errors, less tolerant of magnitude errors
3. **Context effects:** High-risk situations accelerate trust loss
4. **Error patterns:** Single large error vs. accumulated small errors

**Deliverables:**

- Simulation results validating model face validity
- Parameter sensitivity analysis
- Identification of which parameters most critically need empirical fitting

### **Phase 3: Pilot Human Subject Study (Months 3-5)**

**Objective:** Fit model parameters to real human trust dynamics data.

**Protocol:**

- N = 8-12 healthy young adults
- Lower-limb prosthesis simulator (able-bodied adapter)
- Measure Big Five personality (BFI-44 questionnaire)
- Robot varies assistance: (a) good, (b) timing errors, (c) performance errors
- Collect: Trust ratings (every 10 steps), physiological (HR, EMG), behavioral (gaze, compliance)
- Motion capture for gait metrics (symmetry, balance)

**Analysis:**

- Fit  $\alpha_{loss}$ ,  $\alpha_{gain}$ , error thresholds via maximum likelihood
- Validate personality predictions: correlate fitted parameters with Big Five scores
- Test model's ability to predict trust trajectories for held-out trials

**Deliverables:**

- Empirically validated trust dynamics model
- Dataset for controller development
- Conference paper (HRI, ICORR, or similar)

### **Phase 4: POMDP Controller Development (Months 6-12)**

**Objective:** Build adaptive controller that infers hidden psych-state and optimizes actions.

**Components:**

1. **Belief State Representation**
  - Hidden state: (trust, personality traits, physical state)
  - Maintain probability distribution via particle filter or Gaussian belief space

## 2. Observation Model

- $P(\text{observations} | \text{hidden state, action})$
- Likelihood evaluation for Bayesian update
- Use validated correlation matrices from Phase 1-3

## 3. Action Space

- Discrete: {low, medium, high} assistance levels
- Continuous: timing offset, impedance parameters
- Start with discrete for tractability

## 4. Reward Function

- Multi-objective: task performance (gait quality) + trust maintenance
- Penalty for trust dropping below threshold
- Trade-off weighting (explore Pareto frontier)

## 5. Planning/Policy

- Online POMDP solver (POMCP, DESPOT, or similar)
- Model-predictive control with belief propagation
- Real-time computational constraints

### Deliverables:

- Functional POMDP controller implementation
- Simulation experiments showing adaptation to different user profiles
- Comparison baselines: fixed assistance, trust-agnostic adaptive control

## Phase 5: Real-World Validation (Months 13-18)

**Objective:** Validate controller with human subjects in prosthesis experiments.

### Experiments:

- $N = 20-30$  subjects (powered by pilot study)
- Within-subject design: POMDP controller vs. baseline controllers
- Measure: task performance, trust evolution, user preference, safety metrics

### Deliverables:

- Full validation study
- Journal paper (IEEE Trans. Robotics, Trans. Neural Systems & Rehabilitation, or similar)
- Open-source release of validated framework

---

## 6. Expected Contributions

---

## 6.1 Scientific Contributions

1. First psychophysically-grounded trust perception model for physical HRI
  - Explicit modeling of perception layer (objective → subjective)
  - Integration of individual differences into perceptual mechanisms
  - Validated against cognitive neuroscience and personality psychology literature
2. Empirically validated personality-modulated trust dynamics
  - Quantitative mapping: Big Five → trust evolution parameters
  - Age effects on temporal error tolerance
  - Asymmetric trust update rules with individual modulation
3. POMDP framework for trust-aware adaptive control
  - Principled approach to hidden psychological state estimation
  - Demonstrates feasibility of real-time belief updates
  - Multi-objective optimization (performance + trust)

## 6.2 Practical Impact

1. Personalized wearable robotics
    - Controllers that adapt to user's personality and state
    - Improved user acceptance and long-term adoption
    - Safer operation via trust-sensitive assistance
  2. Design guidelines
    - Identification of which robot behaviors are most trust-critical
    - Personality-specific interaction strategies
    - Age-appropriate assistance paradigms
  3. Generalizable framework
    - Applicable beyond prosthetics: exoskeletons, rehabilitation robots, assistive devices
    - Extensible to other domains: surgical robots, autonomous vehicles, collaborative manufacturing
- 

## 7. Technical Risks and Mitigation

### 7.1 Model Complexity vs. Tractability

Risk: Continuous hidden state space makes POMDP intractable for real-time control.

#### **Mitigation:**

- Start with discretized trust levels (e.g., 5 bins: very low, low, medium, high, very high)
- Use particle filters with adaptive resampling
- Investigate Gaussian belief space approximations
- Hierarchical decomposition: slow personality estimation + fast trust tracking

## **7.2 Parameter Identifiability**

**Risk:** Too many free parameters to reliably fit from limited human subject data.

#### **Mitigation:**

- Fix personality-to-perception mappings from literature (not free parameters)
- Focus empirical fitting on trust dynamics ( $\alpha_{\text{loss}}$ ,  $\alpha_{\text{gain}}$ , thresholds)
- Use informative priors from pilot study
- Regularization to prevent overfitting

## **7.3 Individual Variability**

**Risk:** High inter-subject variability makes personalization difficult.

#### **Mitigation:**

- Hierarchical Bayesian models: population + individual parameters
- Online adaptation: controller refines user model during interaction
- Personality assessment as prior; behavior as likelihood
- Graceful degradation: safe default behavior when uncertain

## **7.4 Real-Time Computational Constraints**

**Risk:** POMDP planning too slow for 100Hz prosthesis control loop.

#### **Mitigation:**

- Two-layer control: slow POMDP (1-10Hz) for high-level strategy, fast PID for low-level tracking
  - Precomputed policy approximations (offline value iteration)
  - GPU acceleration for particle filter updates
  - Anytime algorithms with early termination
- 

## **8. Publications Timeline**

#### **Year 1:**

- Conference paper (HRI, ICORR): "Personality-Modulated Trust Perception in Lower-Limb Prosthetics: A Psychophysical Model"

#### Year 2:

- Conference paper (ICRA, IROS): "POMDP-Based Trust-Aware Adaptive Control for Wearable Robotics"
- Journal paper (IEEE Trans. Human-Machine Systems): "Individual Differences in Trust Dynamics During Physical Human-Robot Interaction"

#### Year 3:

- Journal paper (IEEE Trans. Robotics or Trans. Neural Systems & Rehabilitation): "Validated Trust-Aware Control Framework for Lower-Limb Prosthetics: From Perception to Adaptation"
- 

## 9. Resources and Collaboration

### 9.1 Current Resources

- Lower-limb prosthesis hardware
- Healthy young adult subject pool
- Basic motion capture and physiological sensors
- Computational infrastructure (Python, simulation environment)

### 9.2 Needed Resources

-  Expanded sensor suite: wireless EMG, HR monitor, eye tracker (for gaze/monitoring)
-  IRB approval for human subject experiments
-  Funding for subject compensation (N=30-40 total across studies)
-  Access to clinical populations (future: amputees, elderly users)

### 9.3 Potential Collaborators

- Personality psychology expert (validate Big Five integration)
  - Clinical prosthetist (ground-truth assessment of gait quality)
  - Control theorist (POMDP optimization, computational efficiency)
- 

## 10. Conclusion

This research addresses a fundamental gap in wearable robotics: the lack of principled, psychologically-grounded models of how trust forms and evolves during physical human-robot interaction. By explicitly modeling the perception layer and integrating individual

differences, this work enables a new class of adaptive controllers that can infer hidden psychological state and personalize assistance strategies accordingly.

The current infrastructure provides a solid foundation—validated perception models, personality integration, and biomechanically relevant action representation. The path forward is clear: empirical validation of trust dynamics, followed by POMDP controller development and real-world testing.

This is not merely a theoretical exercise. With access to prosthetic hardware and human subjects, this work can produce empirically validated, deployable technology that meaningfully improves user experience and safety in wearable robotics.

**The ultimate goal:** Robots that don't just physically assist humans, but understand and adapt to their psychological states, building and maintaining trust through personalized, context-aware interaction.