

The Good (Surprisingly Solid)

1. The Perception Model is Well-Grounded

- `perceived_error()` function with Weber scaling, temporal kernels, and context-dependent processing is legit
- Sharpness parameter (**beta**) derived from Big Five is creative and reasonably justified
- Age-based temporal tolerance (**t0**) is empirically defensible
- Distinction between *event* vs *rhythmic* contexts shows understanding of sensorimotor literature

2. Individual Differences are Concrete

- Big Five → observation likelihood matrix is a reasonable operationalization
- Big Five → perceptual sharpness mapping with confidence weights shows nuance
- Not just hand-waving "personality matters"

3. Action Representation has Biomechanical Grounding

- Walking symmetry, balance (**phi-bonacci index**), metabolic cost are actual metrics used in exoskeleton research
- Weighted combination into `performance_error` is reasonable

What You Actually Have

A sophisticated forward model of human trust perception

My Recommendation

You have a solid foundation but grossly oversold what you've built.

The perception model is good research. **Don't call it a POMDP controller when it's not.**

Fastest Path to Publication

1. Finish the trust dynamics (implement the update equation)

2. Create simulation scenarios with different robot behaviors
 3. Show how different personalities respond differently
 4. Validate predictions with pilot human study (**N=20–30**)
 5. Submit as "*Personality-Modulated Trust Dynamics Model for Wearable Robotics*" to HRI or similar
-

Pilot Data Requirements

You won't know the right constants (**alpha_loss**, **alpha_gain**, **threshold**) without pilot data.

Steps:

- Run preliminary experiments with 5–10 subjects
 - Have them rate trust after each robot action
 - Fit parameters to their actual trust trajectories
 - THEN validate the model on new subjects
-

Lower-Limb Prosthesis is Actually Perfect

You said you have healthy young subjects and a lower-limb prosthesis. This is ideal.

Study Design

- Subjects walk with a powered ankle-foot prosthesis simulator (able-bodied adapter)
- Robot varies assistance timing/magnitude
- You measure: EMG (co-contraction), heart rate, gaze, compliance
- Subjects report trust every N steps

Metrics Map Directly







- **Walking symmetry**: measurable with motion capture
- **Balance**: center of pressure, step width
- **Metabolic cost**: indirect calorimetry or EMG-based estimates

Psych-State You Can Observe

- **Trust**: self-report (Likert scale, validated HRI trust scales)
 - **Anxiety**: HR variability, palm sweat (if sensors available)
 - **Behavioral**: hesitation (step latency), reduced compliance (less weight on prosthesis side)
-

Infrastructure: What's Missing

What You Have

-  Perception model (`perceived_error`)
 -  Individual differences (Big Five → beta, observations)
 -  Action representation (performance + timing)
 -  Trust update dynamics (missing piece)
 -  Robot's belief update (POMDP inference)
 -  Action selection policy (POMDP planning)
-

Honest Path Forward

Phase 1: Validate Forward Model

- Implement placeholder trust update (use asymmetric model above)
- Simulate 5 personalities × 3 robot behaviors (good, mediocre, unpredictable)
- Generate synthetic trust trajectories
- Sanity check: Does high-N lose trust faster? Does high-A recover faster?

Phase 2: Pilot Study (N=8–10)

- Measure Big Five for subjects
- Run prosthesis walking with varied robot quality
- Collect trust reports + physiological signals + behavior
- Fit model parameters to actual data
- Check if personality predictions hold

Phase 3: Build the POMDP

- Validated trust dynamics
 - Robot's belief state: $P(\text{trust, traits} \mid \text{observations})$
 - Action space: {`assistance_level`, `timing_offset`}
 - Observation space: {behavioral, physiological} measurements
 - Reward: task performance + trust above threshold
-

Key Questions Before Talking to Advisor

- What trust scale are you using? Likert 1–7? 0–10? Custom?
 - How often do subjects report trust? Every 10 steps? Continuous slider?
 - What sensors do you actually have? EMG? HR monitor? Motion capture?
 - What can the prosthesis control? Assistance torque? Timing? Impedance?
 - What's your IRB timeline? Can you run pilots in 2–3 months?
-

Good News

- Solid theoretical foundation
 - Reasonable infrastructure for the forward model
-

Other Factors Affecting Trust Dynamics

From empirical HRI literature, trust is influenced by:

1. **Transparency/Predictability** (partially via timing_delay)
 - Sudden changes more damaging than consistent poor performance
 2. **Perceived Risk/Stakes** (missing)
 - Walking with prosthesis = high stakes (fear of falling)
 - Trust loss accelerates when consequences are severe
 3. **User's Current State** (partially included)
 - Fatigue reduces tolerance for errors
 - Cognitive load affects perception and trust updating
 - Anxiety_level exists but unused
 4. **Cumulative History** (missing)
 - Error streak vs isolated error
 - Trust has momentum → harder to rebuild after repeated failures
 5. **Locus of Attribution** (missing)
 - Self-blame vs robot-blame influences trust loss trajectory
-

Steven's Law vs Other Models

Steven's Law ($\psi = k \cdot \phi^n$)

- Good for steady-state perceptual magnitude
- Weak evidence strength (4/10)
- Doesn't capture temporal dynamics, context-dependence, or asymmetry

Fisher Information

- Good for optimal estimation, precision of inference
- Wrong application: estimates trust, not how trust changes
- Useful only later for sensor optimization in POMDP

Three-Layer Model ($T = T_{\text{dispositional}} + T_{\text{situational}} + T_{\text{learned}}$)

- Perfect fit for current framework
 - Directly maps to personality traits, context, and learning dynamics
-

Should You Use RL?

No.

- RL = trial-and-error policy learning
 - You're modeling human trust dynamics, not robot learning
 - RL relevant later for robot planning, not trust modeling
-

Recommendation

Use the **Three-Layer Model** because:

- ☒ Empirically validated in HRI
- ☒ Maps directly to personality traits (dispositional)
- ☒ Handles context (situational)
- ☒ Captures learning dynamics (learned)
- ☒ Simple enough to fit with pilot data
- ☒ Interpretable for reviewers

Don't use:

- Steven's Law (not designed for this)

- Fisher Information (wrong stage)
 - RL for trust dynamics (wrong tool)
-

Next Steps

1. Implement the three-layer model this week
2. Run simulations with different personalities + robot behaviors
3. Show model predicts intuitive patterns (e.g., high-N loses trust faster)
4. Start pilot study to fit **α_{loss}** , **α_{gain}** parameters from real data
5. THEN worry about POMDP