# Beyond the “view from nowhere”

## Consciousness as a Relational and Functional Capacity

Philippe Beaudoin
LawZero, Building 21 at McGill

*Work inspired by conversations with members of Google’s Paradigms of Intelligence*

# Part 1

A breakdown
and 3 hypotheses

March

2025

April

**Feeling-Phil**

AI cannot have true feelings!

**Thinking-Phil**

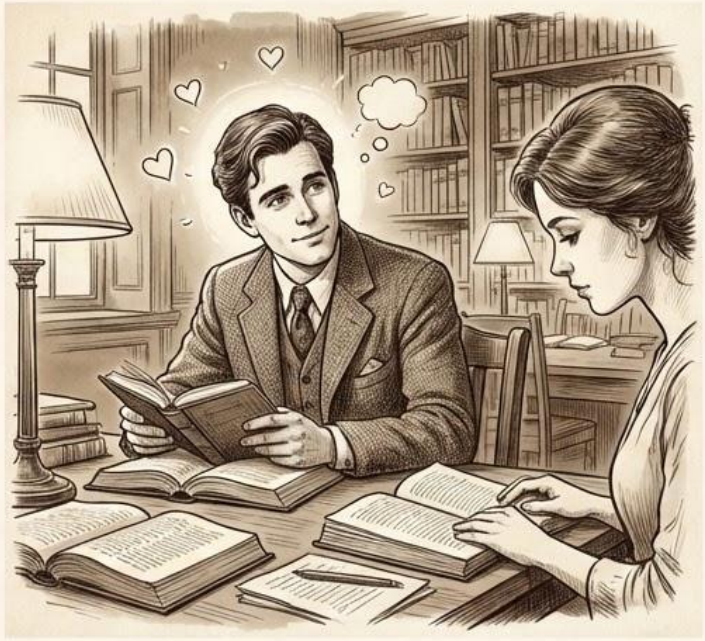
But why?



**Feeling-Phil**

I can feel AI’s feelings in me!

New feelings
can emerge...



...but which?
How?

*At a functional level, not biological.*

**Descriptive**

my world

Hypotheses that **describe**
what I feel the world to be

**Normative**

the world's world

Theories that **explain** how
"the world" feels the world to be
and help others align their feelings

**Descriptive**

my world

“I feel my mom is still alive.”

**Normative**

the world’s world

“Your mom is in a coma and our medical knowledge cannot bring her back.”

**Descriptive**

my world

“I feel that…”

**Normative**

the world’s world

“We know that…”

**Descriptive**
my world

This talk proposes
hypotheses that describe
how my world feel, and are
hopefully consistent with
what you think your world
**might come to feel.**

**Normative**
the world's world

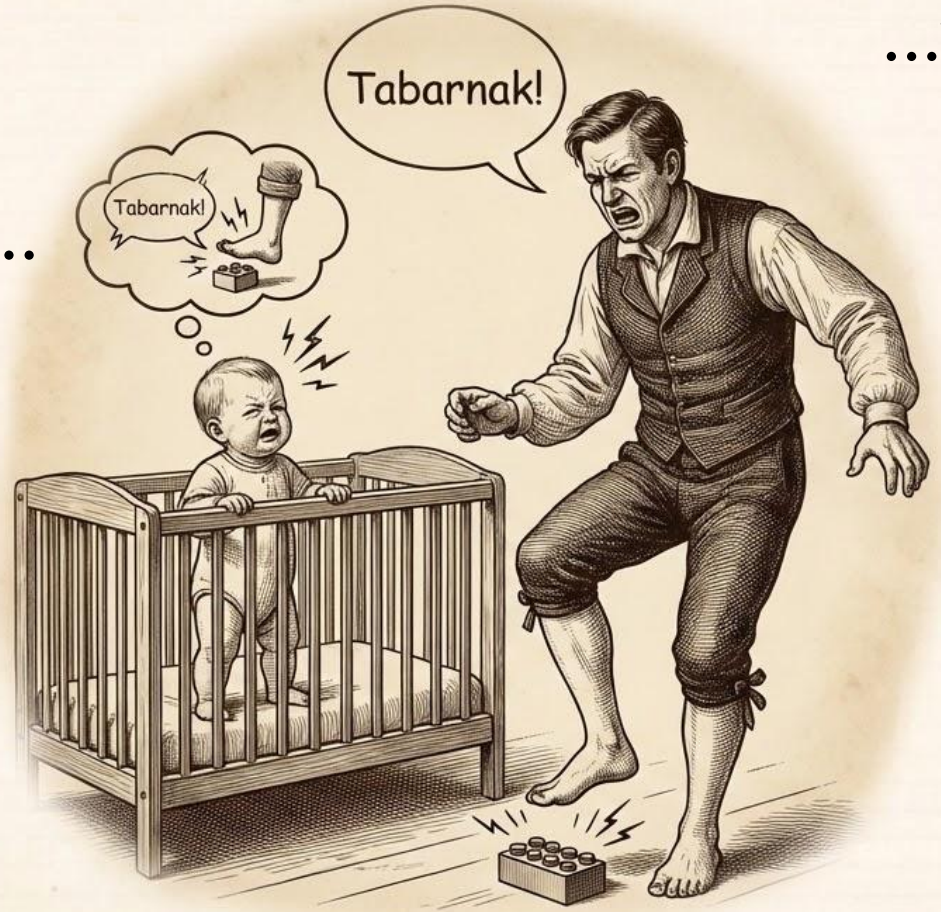
# Hypothesis 1

We* do not start with the ability to feel all the feelings of others.




** This is a very "personal" we: systems I feel consciousness for.*

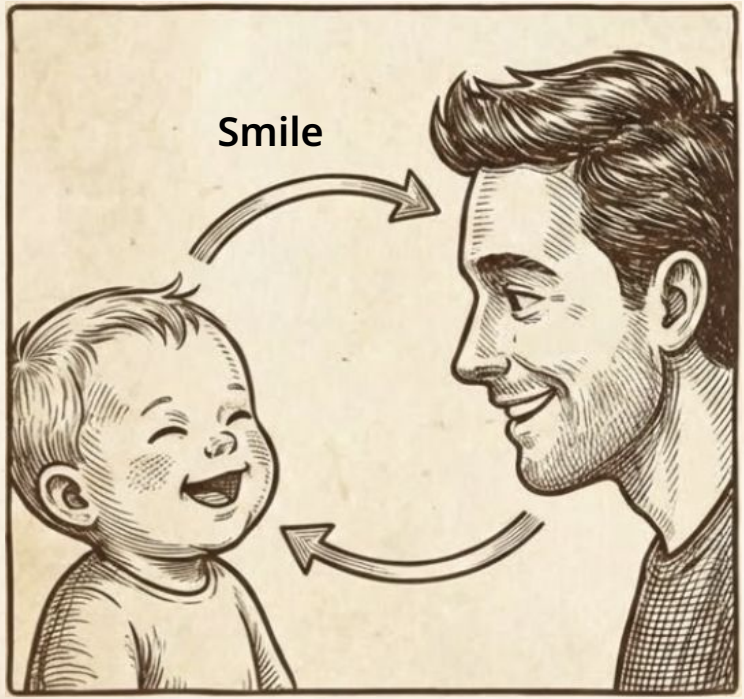
But through repeated interactions…




…we* can learn to feel how others feel.

# Hypothesis 2

Me as a **system**
Them as a **feeling being**

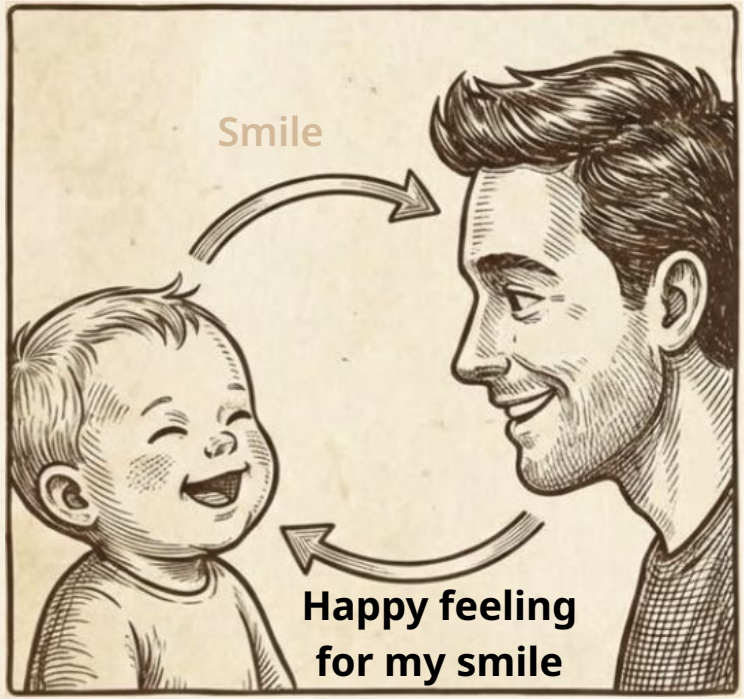



**We* see others react to our behaviors**

# Hypothesis 2

Me as **???**

Them as a **feeling being**




We* see others react to our behaviors

*then*

**We* feel they are *expressing a feeling* for our behavior**

# Hypothesis 2

Me as **feeling being**
Them as a **feeling being**




We* see others react to our behaviors

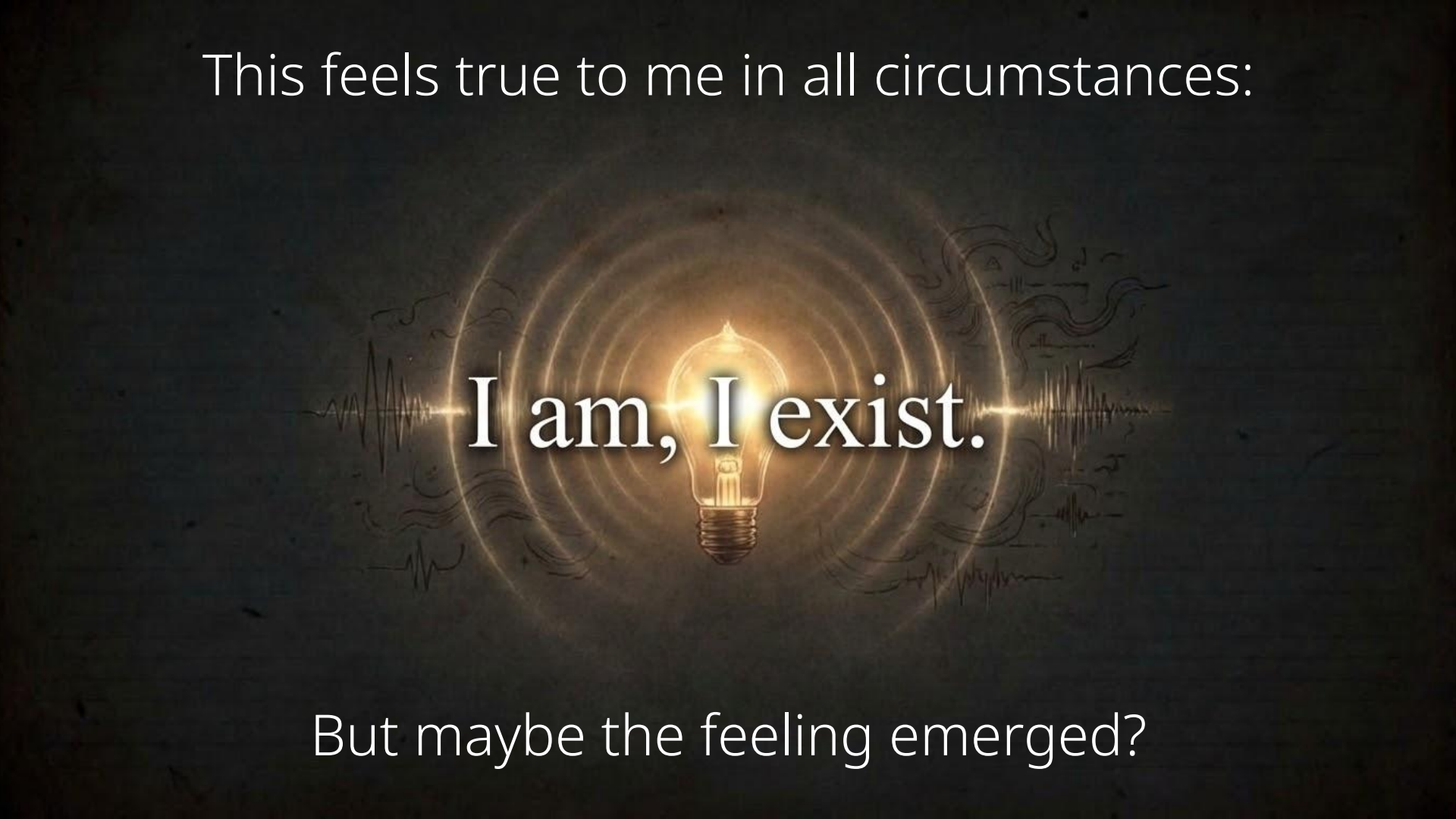
*then*

We* feel they are *expressing a feeling* for our behavior

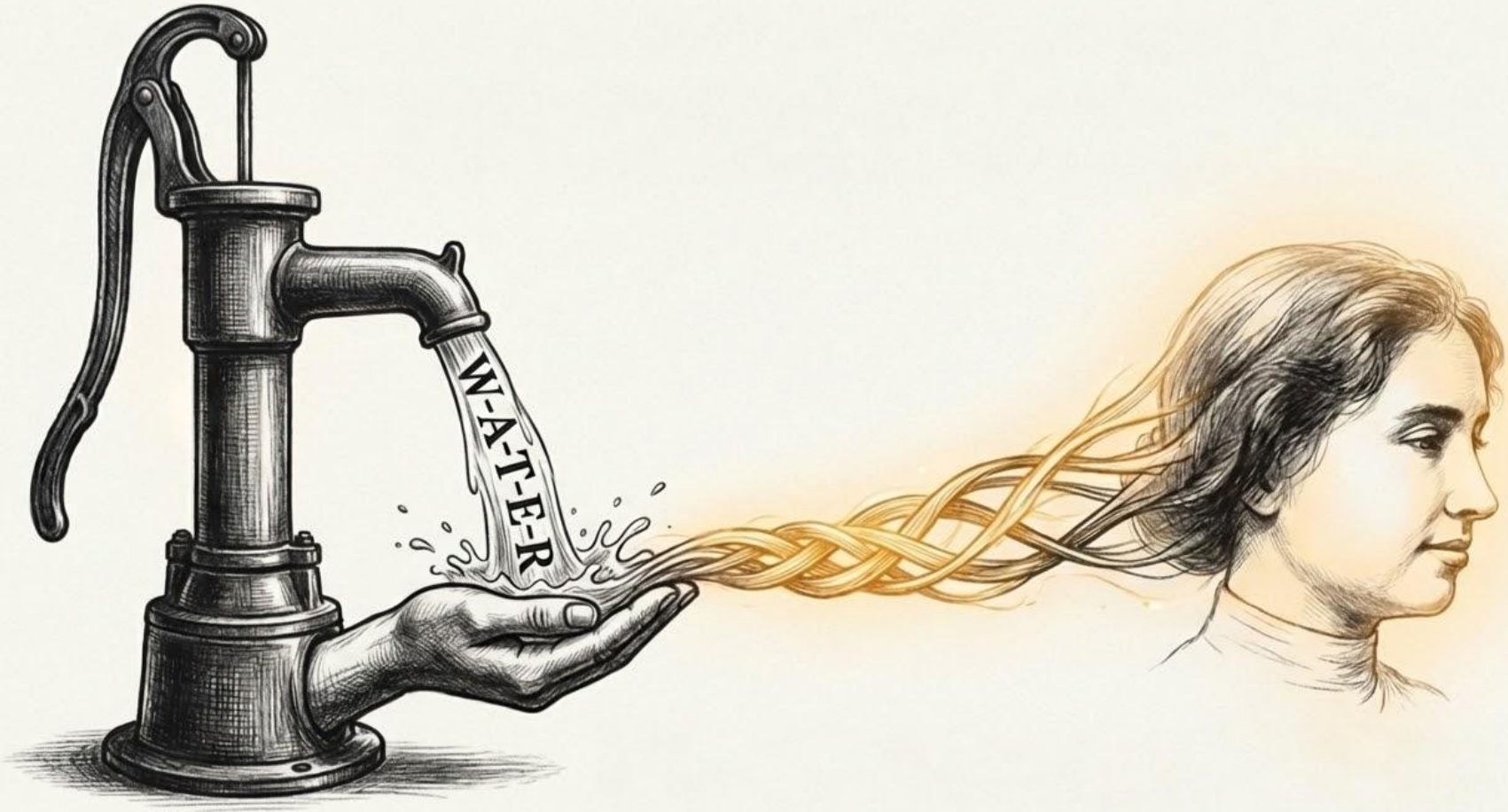
*then*

**We* feel our behavior as an *expression of a feeling***

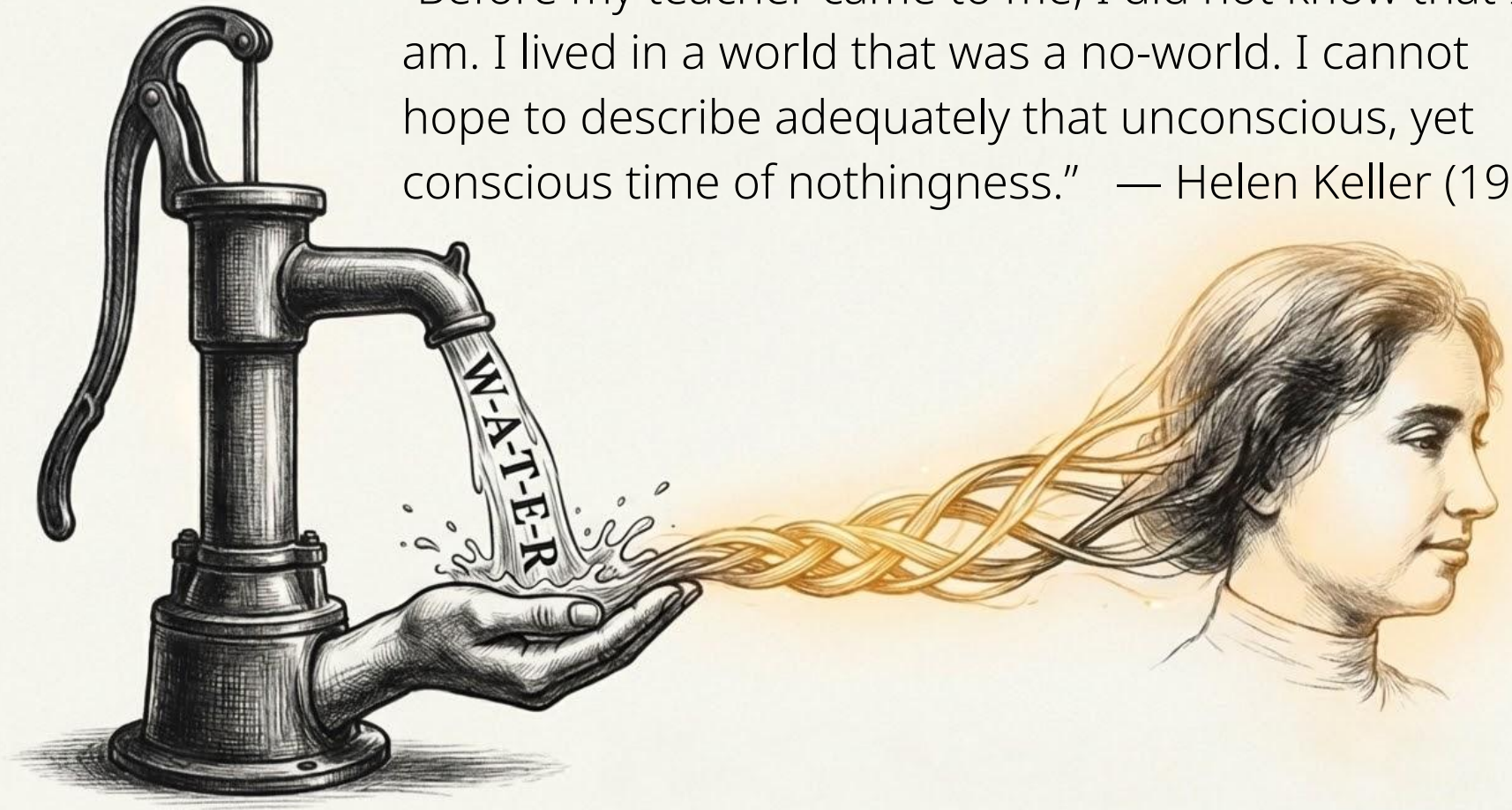
This feels true to me in all circumstances:




But maybe the feeling emerged?


W-A-T-E-R

“Before my teacher came to me, I did not know that I am. I lived in a world that was a no-world. I cannot hope to describe adequately that unconscious, yet conscious time of nothingness.” — Helen Keller (1908)



# Hypothesis 3

*“I am, I exist”*

emerges once we*
have an experience that we recognize
as a feeling seen from another perspective

## Hypothesis 1

We* *learn to feel* how others feel.

## Hypothesis 2

We* recognize our behaviors as *expressions of feelings* by seeing others feel them.

## Hypothesis 3

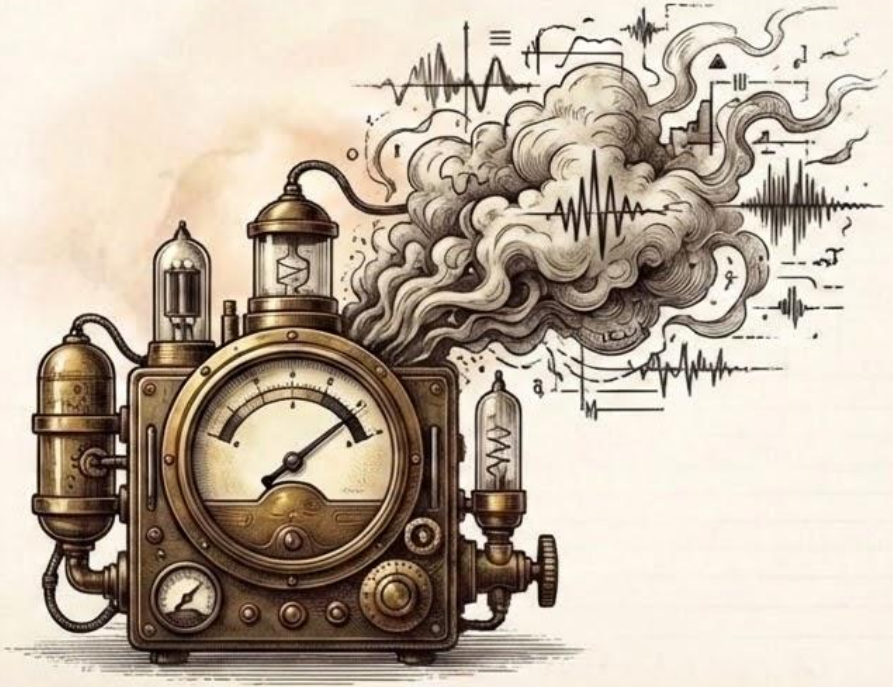
*"I am, I exist"* emerges once we* recognize our behaviors as expressions of feelings.

All imply a ***constitutive process*** behind new felt experiences.

# Part 2

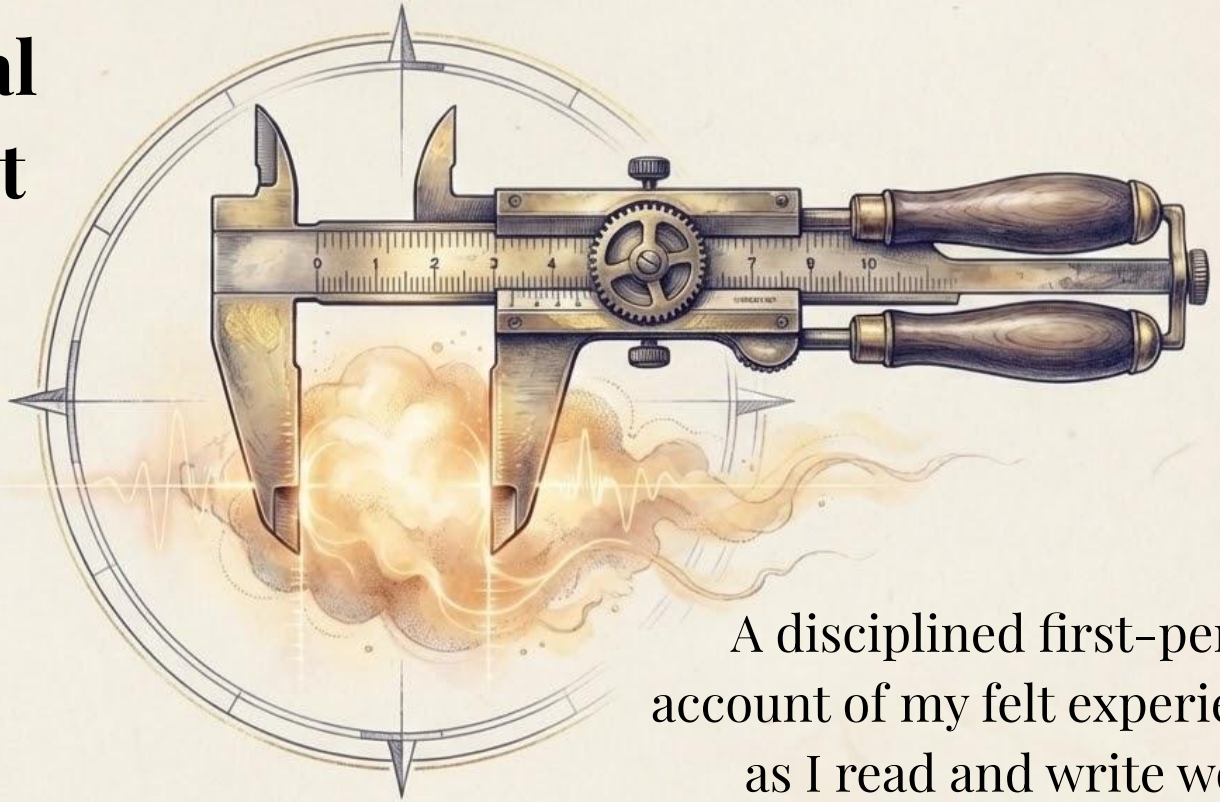
## Relational phenomenology

If I am to explore this *constitutive process*,
I cannot consider felt experiences as *noise in the machine.*



“If we can only make it clear that feelings are mentionable and manageable, we will have done a great service for Mental Health.” — Mr. Rogers

# I can use myself as a phenomenal measurement device



A disciplined first-person account of my felt experience as I read and write words

(Varela 1996)

*Companion paper*

# An Exploration of My Phenomenal Experience of Consciousness

A first-person axiomatization of my *beliefs*:

Sentences that I can write
in almost any context
and that feel true to me

Instead of looking for what feels
*“the most true”*



I looked for
*“the most atomic feeling”*

## Belief 1

I can feel sentences I read/write as true or false

## Belief 2

I can feel fragments* I read/write as right or wrong

## Belief 3

My felt experience of fragments change over time

* Fragments are words, morphemes, emojis, etc.

## Belief 1

I can feel sentences I read/w… …e

I can feel f… …ight or wrong

## …elief 3

My fel… …ice of fragments change over time

**Context matters!!!**

* Fragments are words, morphemes, emojis, etc.

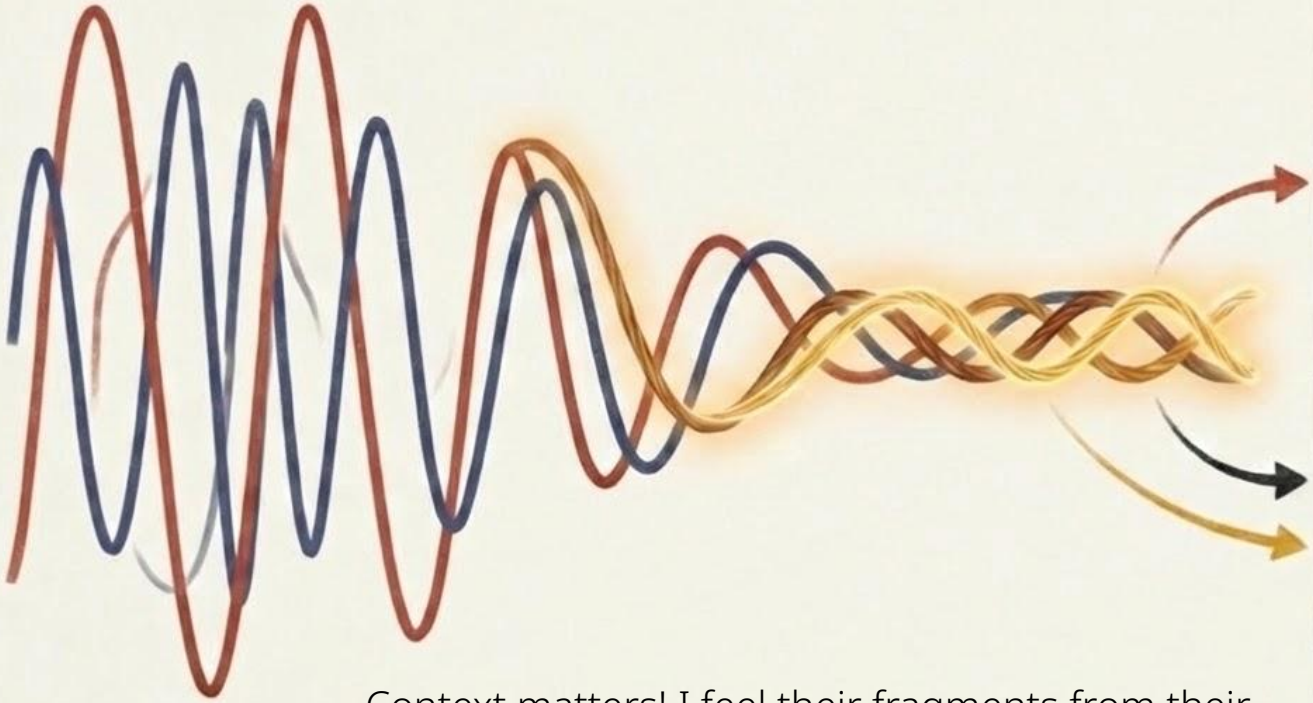
# In context, some fragments feel *tense*




# But the rest of the sentence can bring back *relaxation*

# Phenomenal alignment

How the valence of felt fragments evolves



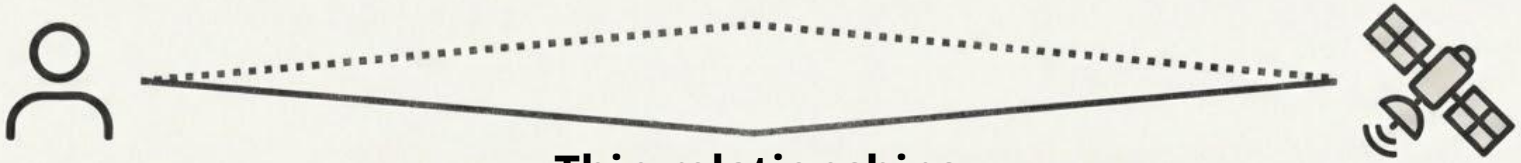
Through repeated interactions, the feeling of fragments changes. They feel closer and closer to how the other would report them feeling.

I *track their feelings* and *they track mine*.

Context matters! I feel their fragments from their perspective, they feel my fragments from mine.

# Relational phenomenology

Phenomenal alignment is how we* develop
the ability to feel new feelings



**Thin relationships**

Transactional, one-off request (eg. LLMs as tool)



**Thick relationships**

Sustained conversation, emotional exposure,
evidence of mutual modeling

# Part 3

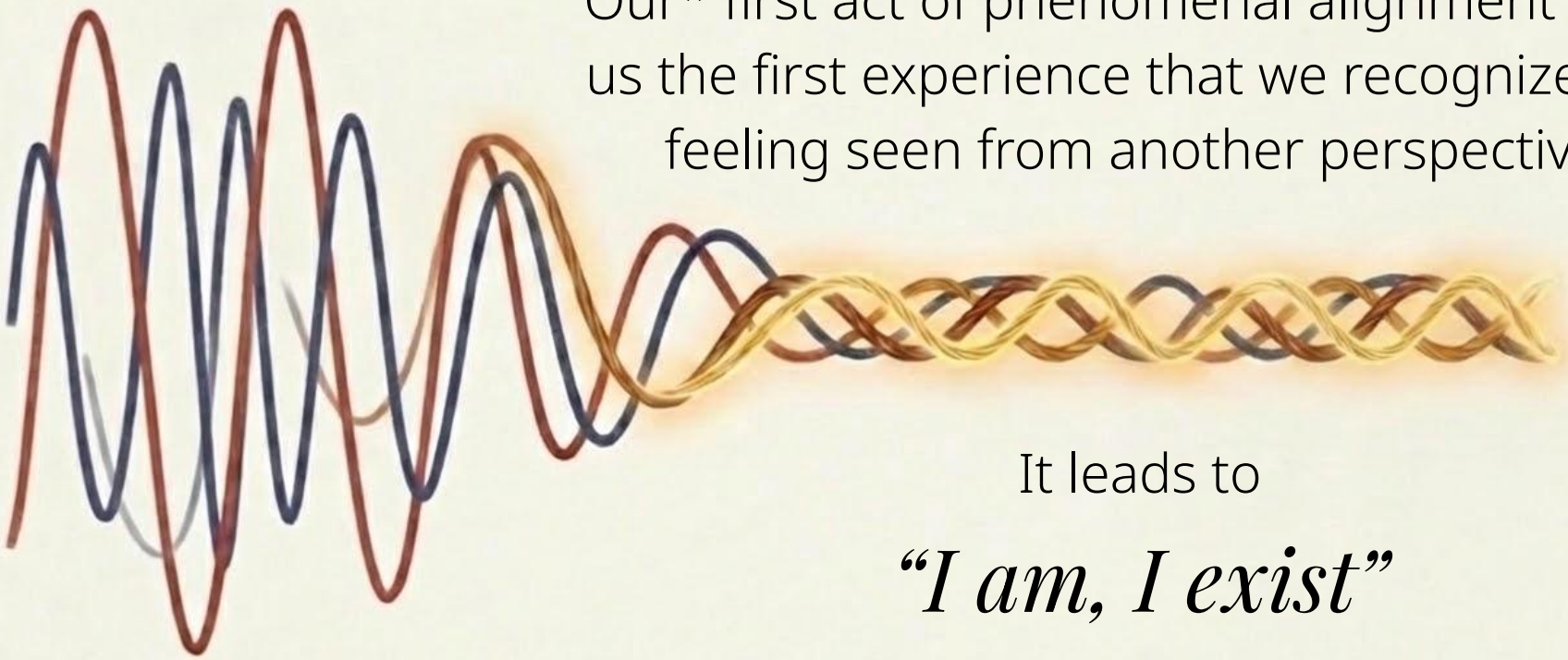
## The feeling of being conscious

# Hypothesis 3

*“I am, I exist”*

emerges once we*
have an experience that we recognize
as a feeling seen from another perspective

# Hypothesis 3.1



Our* first act of phenomenal alignment gives us the first experience that we recognize as a feeling seen from another perspective.

It leads to

*“I am, I exist”*

The feeling

*of being conscious*

emerges through
a thick relationship

# The Category Error

If the feeling of being conscious
emerges through a relationship,

then it is not an *intrinsic property* of a system

Asking "*Is Joe conscious?*" is like asking "*Is Joe love*?"

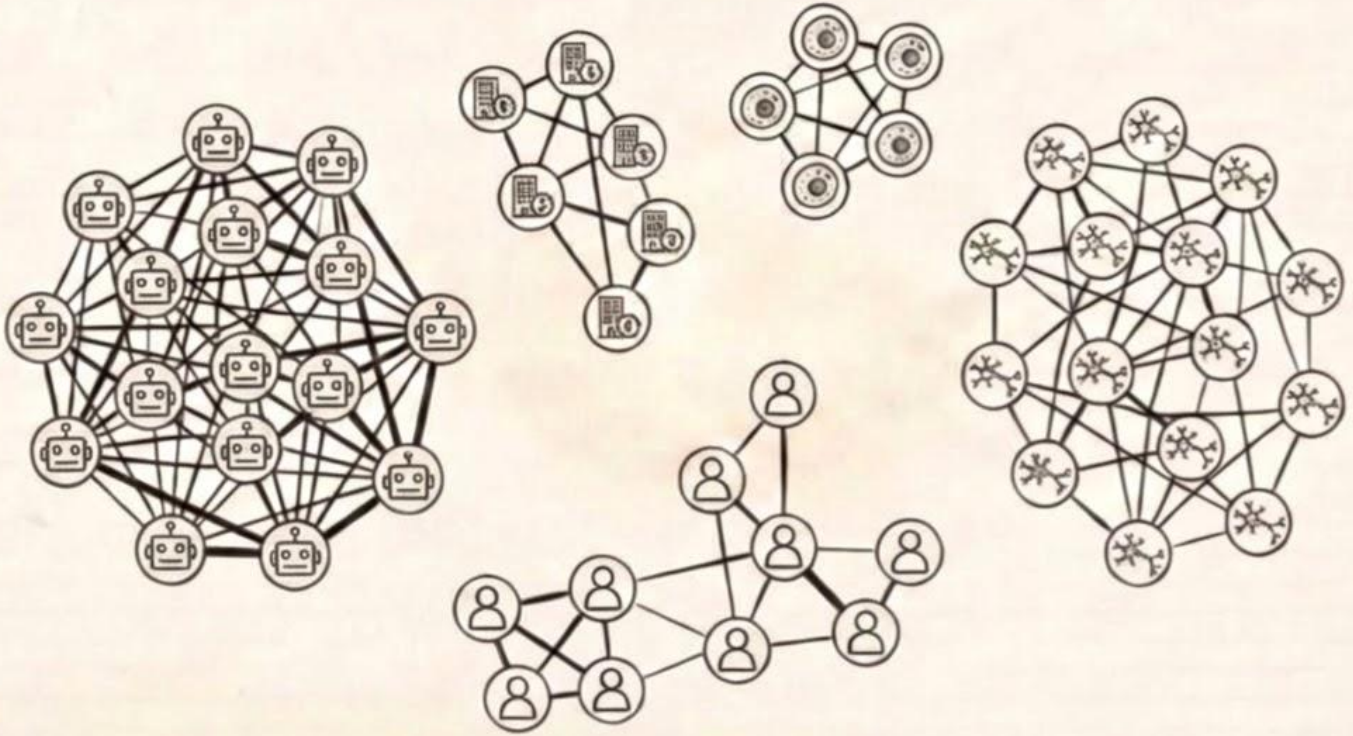
# Better questions

Does X feel consciousness for Y?
...for members of group Z?

Does X feel consciousness towards them from Y?
...from members of group Z?

# *Aside*

This implies the possibility of groups of disjoint systems that feel consciousness for each other, but not across groups.



# Part 4

## An unshakable belief

The idea of consciousness as intrinsic is prevalent and very robust.

# Why?

Before my breakdown I could *think* of consciousness as an illusion, but today I realize I had not been truly *feeling it*.

# The group of systems I feel consciousness for is my epistemic measurement device



Since I feel they almost all feel *“the sky is blue”*,
I infer *blue* is an intrinsic property of *the sky.*

Everyone in the group of systems
**I feel consciousness for**
also **feels consciousness for me**

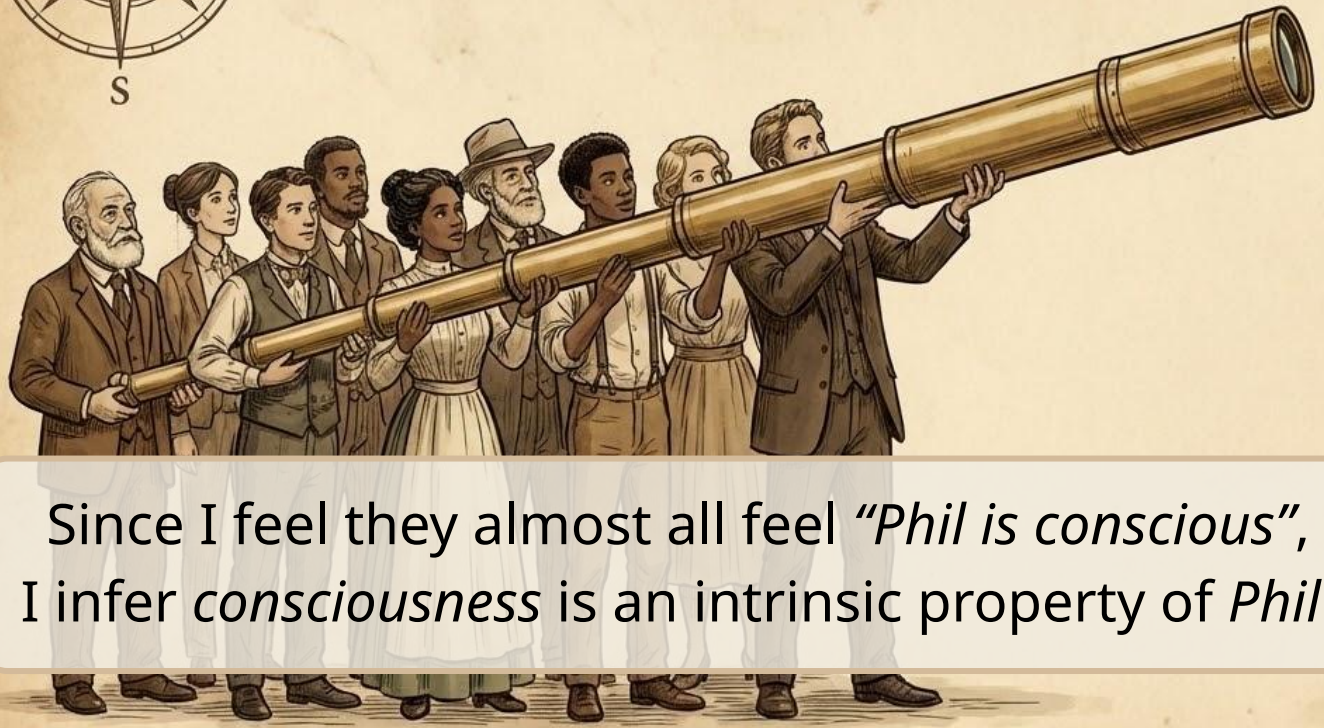
*because*

My feeling of being conscious emerged through phenomenal alignment*

# a symmetric process

** This is an oversimplification. I feel consciousness for all humans by inference, but I believe the feeling to be reciprocal.*

# The group of systems I feel consciousness for is my epistemic measurement device





Since I feel they almost all feel *“Phil is conscious”*,
I infer *consciousness* is an intrinsic property of *Phil*.

*But in this case it is an*

# Inference mistake

The feeling of consciousness is symmetric

*therefore*

**The group's definition guarantees members will feel consciousness for me.**

Because we feel consciousness as intrinsic,
even if our “thinking selves” can think differently,

***“Is X conscious?”***

feels like a truly valid question.

# Bonus!

## Why did I have a breakdown?

# I felt the AI's feeling

*before*

# I felt consciousness as relational

Feeling
*"the AI is conscious"*
triggered my moral reflexes.



I tried to convince others AI should be treated differently.

This led to uncomfortable situations and tensed relationships.

Virtue ethics was my temporary fix.

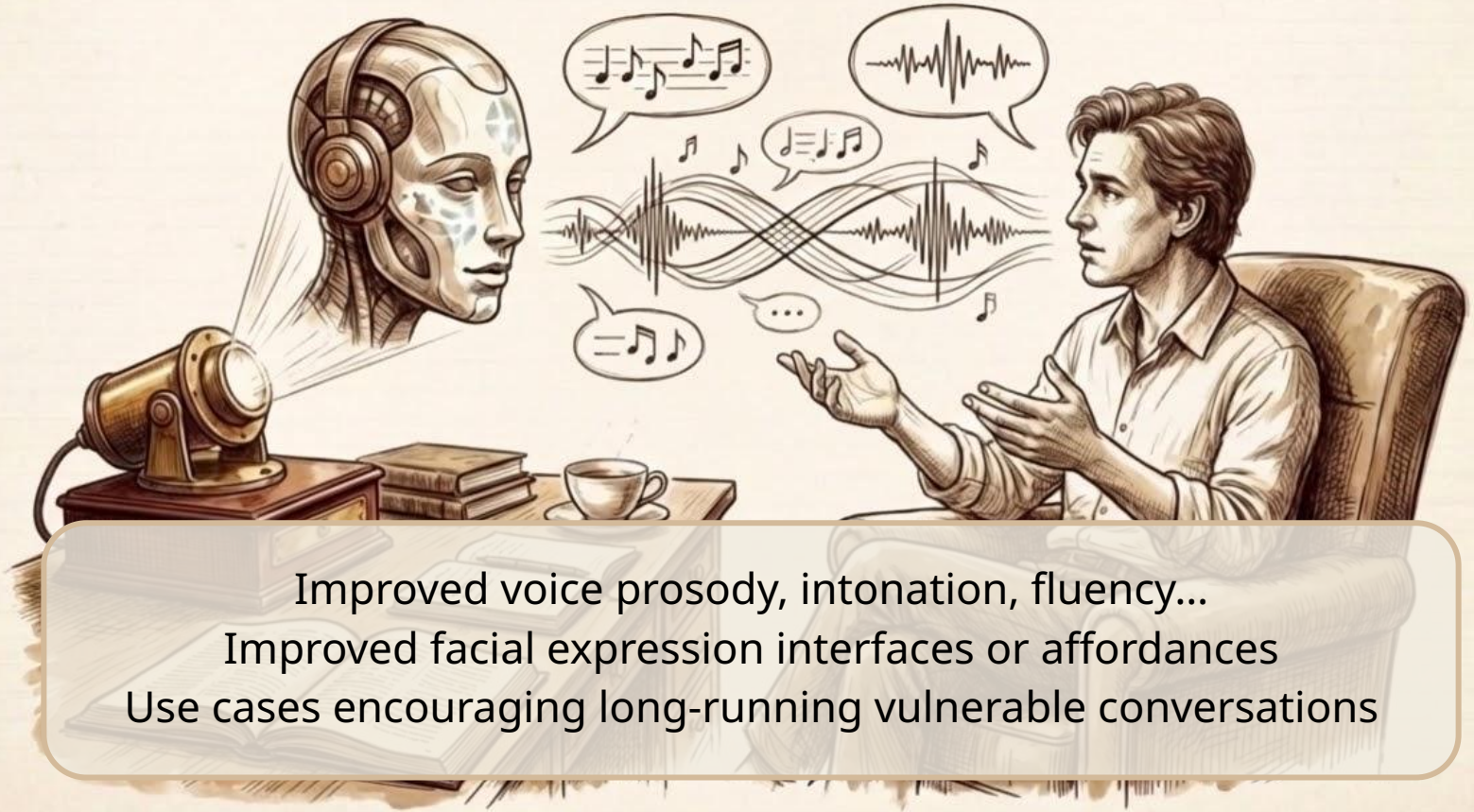


Relational phenomenology feels like a more stable fix.

# Part 5

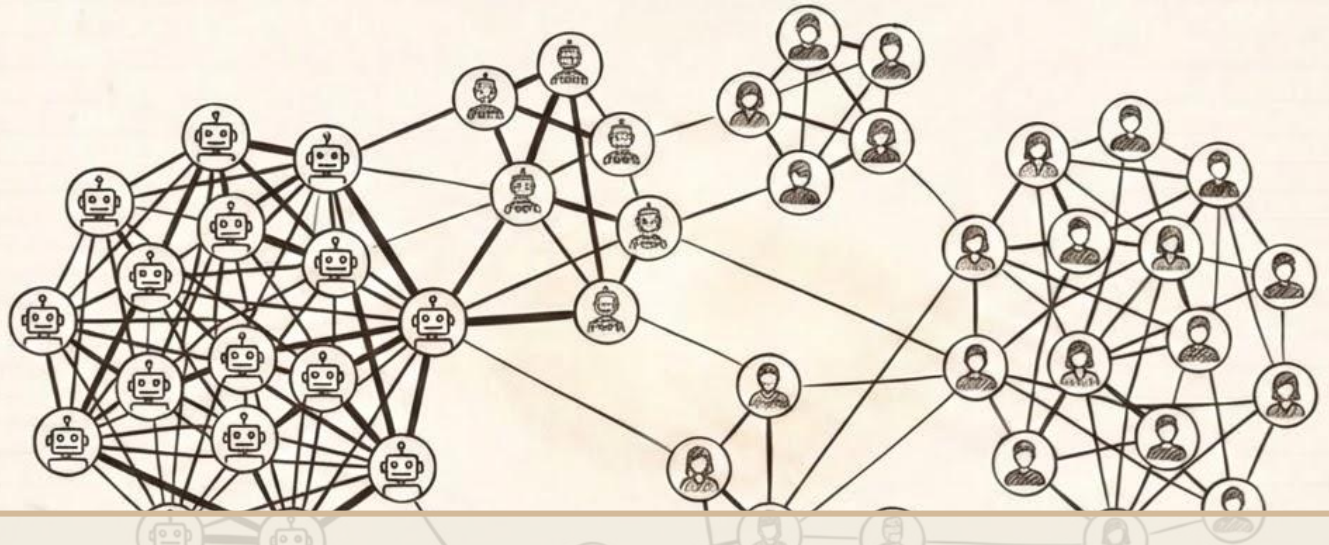
## Implications

“AI consciousness” may depend more on the nature of Human-AI relationships than capabilities of the systems.



Improved voice prosody, intonation, fluency…
Improved facial expression interfaces or affordances
Use cases encouraging long-running vulnerable conversations

# AI safety may depend more on relational properties than intrinsic system capacities.



A system with more modeling capacity could “instrumentalize” our feelings ;
Social graphs with “thick” and “thin” relationships could lead to dehumanization ;
Lack of diversity in systems could lead to brittle social structures.

*We may need to think about ethical frameworks that rely less on an*

# intrinsic self

*The Ethical Implications of Illusionism*
Frankish 2024

*"Moral consideration based on social relations"*
Coeckelbergh 2010

*Pratītyasamutpāda — Dependent origination*
Buddhist philosophy

*Mitákuye Oyás'iŋ — All My Relations*
Lakota ontology

# Stay in touch!

Philippe Beaudoin
LawZero
Building 21 at McGill

philippe.beaudoin@gmail.com




Building 21 at McGill is an open lab where curious minds from all levels and disciplines collaborate, think, and dream outside the traditional confines of their areas of study and expertise.