

Two Hard Problems

A discussion of related problems in ethics and consciousness

Motivations

1. My belief that consciousness and ethics are crucially intertwined.
2. My fear that humans “invented” value.
3. My suspicion that language is requisite for consciousness.
4. My view that what makes consciousness unique is its relationship to verification, i.e. an epistemic property.

Why are these hard problems?

The Hard Problem of Consciousness

How can the tools of science possibly explain subjectivity, i.e., first person qualitative experience?

The Hard Problem of Meta-Ethics (Nihilism)

How can ethics be grounded when religion (no benevolent god), logic (is \neq ought), and evolution (might makes right) cannot be its foundation?

Moral Patients and Moral Agents

- A moral patient is a something that can be helped or harmed but cannot be held responsible for its actions.
- Humans are moral agents. Moral agents can be held responsible for their actions. All moral agents are also moral patients.

Assertion: most people believe there are exist *some* moral patients.

Question: what is that makes something a moral patient?

The Presentation

1. Most folk ethical views rely on consciousness to secure the value of moral patients.
2. These ascriptions are obscured by the partial and anthropomorphic (propositional) nature of empathy.
3. This relationship contributes to the perception that the hard problem of consciousness is intractable.
4. Higher-order theories of consciousness present an intuitive solution to this puzzle.
5. Strict naturalistic ethical reductionism can stretch our moral intuitions to include all non-human animals and AI without the need for consciousness.

How do we understand the value of moral patients?

Moral agents express our own desires in language and understand the value of others desires using empathy. We can use these mental models to resolve moral disagreement and seek opt-in dynamics using frameworks such as utilitarianism, deontology, and contractualism.

For moral patients, we must make inferences from their behavior to their values and internal states.



***Unitree's
First General-purpose
Humanoid Robot***

*the preview of
half-a-year achievement*

The anatomy of (my) consciousness ascriptions

1. A display of behavior characteristic to a familiar conscious agent (pain, love, grief, etc.)
2. An instance of an affective (emotional) empathetic resonance on the part of a conscious agent
3. An attempt at perspective taking (cognitive empathy)
4. A dissection of the mechanism to understand the source of its internal states
5. A confirmation or disconfirmation of the “appropriateness” of the empathetic response, i.e. a consciousness ascription

Some potential entities with moral status

- An agent that is inclined to accept the hard problem of consciousness when prompted appropriately
- An agent that develops future plans based on counterfactual worlds
- An animal that cries in isolation after the loss of its child
- An organism that influences other organisms by pointing
- A program that gets offended
- A plant that shys from the sun
- A large rock that resists the flow of a river
- An ion that maintains its charge

Empathy is not sufficient for consciousness

The hard problem is *partially* the result of the misuse of empathy in consciousness ascriptions and concerns over the loss of value for cognitively dissimilar agents.

- Affective empathy relies on the assumption that your behaviors indicate your internal states.
- Cognitive empathy relies on the assumption that your internal states resemble mine.

This confusion contributes to the belief that consciousness must be atomic, intrinsic, ineffable, and private.

This is not to say there is no relationship between the two!

What is consciousness then?

- The hard problem arises from a difference in access between our semantic and perceptual knowledge.
- Our internal states are simulations, and we we always have some “residual” that we will call atomic, ineffable, intrinsic, private, etc.
- We think that this residual perceptual information can be isolated, but it only exists within a very particular representational context.

This is confusing because the higher-order representations are necessary for consciousness, but the first-order representations are still what matters.

Why can't we just have the residual?

Ned Block provides the following dichotomy for thinking about consciousness:

- P-Consciousness – an atomic quality of perception or awareness (phenomenality)
- A-Consciousness – an introspective, propositional state of “noticing” p-consciousness (access)

What would p-consciousness without a-consciousness look like?

What's left for our moral patients?

Experiential value and empathy are the foundation of human morality. They are what lead us to care about ourselves and others. But these are complex forms of more basic phenomena of **valence** and **resonance**.

- In physics, valence and resonance are tools for generating complexity and coordination.

We can arrive at a universalist ethics by viewing biology as an outgrowth of persistent valence and resonance, and all moral patients along that spectrum.

Appendix

Valence and Value

- Valence (charge) is a physical mechanism for generating complexity and interaction
- Biology creates composite valence over time (minimal self-interest)
- Individual humans convert valence to value through explicit goals

Resonance and Empathy

- Resonance is a mechanism in physics for solving coordination problems
- Biology create weakly emergent behavior through resonance
- Resonance for human beings is mediated by beliefs and *values*
- Our empathy is unconsciously bounded by a multitude of characteristics

Valence/Value + Resonance/Empathy = Morality

OR

Ethics is the [intuitive and deliberative considerations] of the [ends of another agent]



Why does this framing of ethics matter?

- A theory of ethics defines the maximum size of a tribe – a broader theory of ethics makes us resilient towards the future and the past.
- This account avoids a descent into nihilism by grounding ethics in our broader physical and informatic context.

The Hard Problem of Consciousness

Can there be a strictly third-person account of the necessary and sufficient conditions for consciousness?

Would it be possible to create an agent that displayed all appropriate behaviors (including judgements/perceptions) but had no subjective internal states?

An useful way to understand philosophical zombies:

Possibility 1: Philosophical zombies are conceivable, but not implementable

Possibility 2: Philosophical zombies are not conceivable

Possibility 3: Philosophical zombies are both conceivable and implementable