The Edge of Thought

The paradoxes of existence and unsolved philosophical questions in psychology

Jochen Fromm

Third Draft Version 0.3.5 2026-09-20

Table of Contents

Image credits

Front cover images from left to right:

Unsplash photo of white and black lighthouse near body of water during night time from user **Red Zeppelin**
https://unsplash.com/photos/white-and-black-lighthouse-near-body-of-water-during-night-time-Z_Jdq_xKwb8

Unsplash photo of woman in cardigan leaning on bookshelf from **Matt Antonioli**
https://unsplash.com/photos/woman-in-cardigan-leaning-on-bookshelf-uvbMz9dB0DQ

Unsplash photo of Trinity College Dublin from user **Jo Lord**
https://unsplash.com/photos/a-large-building-with-many-windows-mhKJcZycew0

Artemis II photo of setting Earth from **NASA**
https://www.nasa.gov/image-detail/amf-art002e009289/

Back cover images from left to right:

Unsplash photo of white rabbit on green grass during daytime from **Guillermo Casales**
https://unsplash.com/photos/white-rabbit-on-green-grass-during-daytime-LQfcolSv2M0

Unslpash photo of green grass field and trees during daytime from **Daniel Stone**
https://unsplash.com/photos/green-grass-field-and-trees-during-daytime-28LH9sPSk98

Unsplash photo of woman in black shirt and brown shorts walking on brown dirt road from **Erin O’Brien**
https://unsplash.com/photos/woman-in-black-shirt-and-brown-shorts-walking-on-brown-dirt-road-during-daytime-1TTvvTxuXLw/

Unsplash photo of woman reading a book from **JoelValve**
https://unsplash.com/photos/woman-reading-book-R6j5GUP-w8Y

Unsplash photo of street under a night sky full of stars from **Evgeni Tcherkasski**
https://unsplash.com/de/fotos/ein-mann-der-mit-einem-fahrrad-eine-strasse-unter-einem-nachthimmel-voller-sterne-hinunterfahrt-T75GioCdT-0

Acknowledgements

The reason why I wrote this book is because I wanted to preserve the ideas that I have written in my blog. Science is the pleasure of finding things out, as Richard Feynman noticed. I like to find things out, and the blog contains many interesting ideas and thoughts about big questions that should not be lost in my opinion.

Therefore, this book mainly contains the ideas I have collected in my blog in the last 16 years, except the content that was included in my last book “Hidden Genes”. And some new ones. Hopefully the ideas will last longer in this form, even if the blog and I will soon be gone. And if nobody will read it, like my last book which was barely read by anyone, then people who are interested have at least the chance to discover it.

There is a wonderful quote often attributed to Chief Crowfoot of the Blackfoot nation where he asks what life is. He said “a little while and I will be gone from among you, whither I cannot tell. From nowhere we come, into nowhere we go. What is life? It is a flash of firefly in the night. It is a breath of a buffalo in the wintertime. It is as the little shadow that runs across the grass and loses itself in the sunset”. This is my flash in the night, so to speak, before I will be gone.

The persons who inspired me to write all this and some of the content in the blog are Nicholas S. Thompson and his friends from the FRIAM group in Santa Fe, New Mexico, USA, including Eric P. Charles, Glen Ropella, Steve Smith, Frank C. Wimberly and David West. Since a single sentence is not enough, I have devoted the whole first chapter to this group.

The AIs who helped me creating this book were the chat bots currently available. I have written the text itself, but during the creation of this book Gemini from Google, Claude from Anthropic and ChatGPT from OpenAI have been used to summarize the books in the references which helped to understand the subject. I have searched for the relevant textbooks from renowned publishers, summarized their content, and tried to put the puzzle pieces together to obtain the bigger picture. Modern LLMs are good at summarizing large amounts of text in seconds, and we have access to more knowledge and more intelligence than any generation before us. They also were useful to analyze the FRIAM list archives.

Gemini has helped to find the work of Robert Nozick too, and Gemini & Claude also helped to create the visualizations of the differential equations: the Hindmarsh-Rose equations for laughter, the three species Hastings-Powell food chain model for self-awareness and the Wilson-Cowan model in combination with Quantum Cognition to model conscious decision-making and humor.

I found inspiration in many good old blogs as well, for example in "3 Quarks Daily" from S. Abbas Raza et al., Maria Popova's blog "The Marginalian", Susan Stepney's blog "A memory less ephemeral", Brian Hayes's blog "bit-player", Cosma Shalizi's notebooks and his blog "Three-Toed Sloth", and John Carlos Baez’s blog “Azimuth”.

The Berlin state library (“Staatsbibliothek Berlin”) provided me with the necessary material and books to write this text. Modern AIs can summarize almost any book in an instant, but it is still necessary to check and validate their responses. Founded in 1661, it is among one of the oldest and largest libraries in Europe. They have more books than you can read in a lifetime, and most of the books mentioned in this text, including the ones from Robert Nozick, Daniel Dennett, Jaegwon Kim, John Searle, Thomas Nagel, Jerome Busemeyer & Peter Bruza, Andrei Khrennikov, Lisa Feldman Barrett and Jaak Panksepp.

Most of all I would of course like to thank my wife Manuela who has been at my side for more than 30 years now and never lost patience when I spent the day at the library again.

Berlin, September 2026

Introduction

This little book is about the paradoxes of existence and the biggest unsolved philosophical questions in psychology. One of the biggest paradoxes of existence has been identified by French philosopher Jean-Paul Sartre (1905-1980). Sartre's paradox of existence, central to his philosophy in "Being and Nothingness", is that human beings are "condemned to be free" (Sartre, 1943). Humans are "thrown" into the world first and must define what they are only after they have come into existence. He calls this "existence precedes essence" and argues that we have by default no predetermined purpose. This creates an absolute freedom that burdens us with total responsibility for creating our own meaning in an absurd world.

Another French philosopher, Albert Camus (1913-1960), goes a step further and argues finding a meaning is not possible in a meaningless universe. The Absurdist Paradox, as defined by Albert Camus in "The Myth of Sisyphus", is the irreconcilable conflict between humanity's desperate desire for meaning and reasonable purpose, and the silence of an indifferent and meaningless universe (Camus, 1955).

Some of the smaller paradoxes of existence are linked to big, unsolved philosophical questions which have no clear yes or no answer.

- Do we have a **free will** or not?
- We can feel the pain of others, and yet we find it to be a very **hard problem** to understand their **subjective experience.** This is called the hard problem of consciousness. Can we solve it?
- We seem to have a **mind** that is different from the **body**, and yet we don't have it, because we are made of physical matter. Is there a ghost in the machine or not?

In philosophy, a paradox can arise if two worlds collide: myth and science, fiction and reality, inside and outside, micro and macro level, etc. Sometimes it can be resolved by changing or shifting our perspective. Two contradictory things can be equally true when they exist within different contexts, perspectives, or dimensions, for instance on different scales or levels of abstraction.

We will have a look at the various paradoxes of existence for the big philosophical questions in psychology. Some things in it will be a little bit speculative and philosophical, but to explore new worlds and uncharted areas you need to have the courage to leave the territory of solid ground, at least for a short time. Andre Gide said “Man cannot discover new oceans unless he has the courage to lose sight of the shore.”

It is not difficult to find the big questions and the hard problems. We only need to look for unsolved paradoxes and problems where philosophers strongly disagree. John Searle for instance argues that there is no "hard problem" of consciousness (Searle, 2004), while David Chalmers who coined the term thinks it is the hardest problem (Chalmers, 1996) and Jaegwon Kim also thinks the qualia of consciousness cannot be described in physical terms alone (Kim, 2005).

It is even controversial which problem is the hardest one. For some philosophers qualia, the subjective experience of what it is like to perceive something, cannot be fully described, while others think we lack a complete explanation of free will or argue there is an explanatory gap for the mind-body problem.

Despite his hands-on approach and his use of ordinary language philosophy, American philosopher Daniel Dennett (1942 – 2024) always dared to tackle the big questions. There is almost no big question that he has not worked on during his long career. I would call this Dennett’s law of philosophy: for every big, fundamental question in philosophy there is at least one Daniel Dennett book.

- The Intentional Stance about intentionality (Dennett, 1987)
- Consciousness Explained about consciousness (Dennett, 1991)
- Darwin’s Dangerous Idea about evolution (Dennett, 1995)
- Freedom Evolves about free will (Dennett, 2003)
- Breaking the spell about religion (Dennett, 2006)

Since I have discussed evolution and religion extensively in my last book, the remaining questions of intentionality, consciousness and free will are left over for this book. Can these deep puzzles and fundamental problems be solved, have they been solved, will they be solved? Let us take a look.

The problem of free will and the hard problem of consciousness have remained unsolved and slightly out of reach so far. They are obviously as much a psychological as a philosophical problem. It makes sense to tackle these hard problems from both sides, from psychology and from philosophy. I believe Robert

Nozick came closest to finding innovative solutions for them. They will remain hard for reasons explained in this book, but Nozick had original ideas how we might solve them, both for free will (Nozick, 1981) and for subjective experience (Nozick, 2001).

We will also try to look at normal things which are hiding in plain sight that might help to understand the biggest philosophical problems and try to use the latest insights from chaos theory and dynamical systems (Feldman, 2019) and Quantum Cognition (Busemeyer, et al., 2012) to help to understand the dynamic behavior of the underlying complex systems.

Main references

+ John Searle, Mind - A Brief Introduction, Oxford University Press, 2004
+ Jaegwon Kim, Physicalism - or Something Near Enough, Princeton University Press, 2005
+ Jaegwon Kim, Philosophy of Mind, Westview Press, 1996
+ David Chalmers, The Conscious Mind: In Search of a Fundamental Theory, Oxford University Press, 1996
+ Daniel Dennett, The Intentional Stance, The MIT Press, 1987
+ Daniel Dennett, Consciousness Explained, Little, Brown and Co., 1991
+ Daniel Dennett, Darwin's Dangerous Idea, Simon & Schuster, 1995
+ Daniel Dennett, Freedom Evolves, Viking Books, 2003
+ Daniel Dennett, Breaking the spell, Viking Books, 2006
+ Albert Camus, The Myth of Sisyphus, Hamish Hamilton, 1955
+ Jean-Paul Sartre, Being and Nothingness, Simon & Schuster, 1943
+ Robert Nozick, Philosophical Explanations, Harvard University Press, 1981
+ Robert Nozick, Invariances - The Structure of the Objective World, Harvard University Press, 2001
+ David P. Feldman, Chaos and Dynamical Systems, Princeton University Press, 2019
+ Jerome Busemeyer and Peter Bruza, Quantum Models of Cognition and Decision, Cambridge University Press, 2012

Where it started

The FRIAM Group and the SFI

The Santa Fe Institute (SFI) is an independent theoretical research institute located in Santa Fe, New Mexico, United States and dedicated to the multidisciplinary study of the fundamental principles of complex adaptive systems, including physical, computational, biological, and social systems. It was founded in 1984 by a group of prominent theoretical scientists, physicists and researchers including George Cowan, Murray Gell-Mann, David Pines and Herb Anderson. Many of them were physicists who had worked at the Los Alamos National Laboratory which is best known for its central role in helping develop the first atomic bomb. Early members include the Nobel Prize winners John Hopfield, Murray Gell-Mann and Philip W. Anderson.

Murray Gell-Mann proposed the quark model that became an essential part of the standard model of particle physics in 1964.1984 was the time when the standard model of physics had already been formulated. Everybody got access to personal computers, and the world wide web appeared. Computers promised new insights by advanced calculations and simulations. Instead of shattering matter into the smallest possible pieces, the question was “how can complex systems emerge from the interaction of simple elements?”. How can agents who interact by following simple rules lead to complex results? How can the wonderful diversity of the universe arise out of a set of simple basic laws and fundamental elements? (Gell-Mann, 1995). To solve this task the SFI founders looked for new connections. They explicitly wanted to break down boundaries between disciplines and bring together people working on problems that crossed the field boundaries.

A few years after the SFI was founded another group emerged in Santa Fe: a small, distributed, informal group that embodied the interdisciplinary ideal associated with the Santa Fe Institute. The FRIAM group is not an institution with programs and funding, but a loosely coupled discussion group in which people from different disciplines and countries exchange ideas and allow unexpected connections to emerge. Although it has no funding, it proved to be remarkably persistent. The group lasted for over 20 years and is still active. Its longevity is itself part of the phenomenon. The combination of mailing list and Friday-morning meetings proved to be successful. Technology changed — from early computing and agent-based modeling to AI and large language models — but many underlying questions stayed the same.

FRIAM is the abbreviation for FRI-day morning (AM). It stands for Friday Morning Applied Complexity Coffee Group. The FRIAM group is just a simple mailing list of people in Santa Fe on Friday mornings to discuss topics they are interested in while having a few cups of coffee. The mailing list was created before the big tech corporations changed the internet as we know it. Before the WWW people discussed in forums and by using emails and mailing lists. Emails existed before the WWW made the internet available for everybody in a browser. A mailing list is an old-fashioned discussion group where people can discuss topics they are interested in. As the name suggests the FRIAM group is interested in complexity and complex systems.

What makes it interesting that it is a rare community that can take the most famously unresolvable questions in human history and reframe them into executable, computational, and observable models - all while drinking morning coffee in Santa Fe. The group itself is a bunch of extremely curious people. Perhaps it became something which the founders of the SFI dreamed of - a place where people from different disciplines and different continents who think outside the box, exchange ideas to explore new interdisciplinary insights, and try to find new connections. Gell-Mann's idea was to bring together people with broad interests allowing connections to develop that had not been planned. FRIAM conversations often work exactly this way.

The physical FRIAM group consists mainly of people in Santa Fe, but the mailing list has attracted people from all over the world living in Australia, South Africa, Germany and India. They are ethologists, mathematicians, programmers, chemical engineers, journalists, AI historians, complexity scientists, software developers and software architects. They were experts in complexity science, mathematics, software engineering, artificial life, ethology, psychology, philosophy, journalism, chemical engineering, AI history, and other fields.

The FRIAM's structural magic lies in its cross-disciplinary friction:

- Psychology, Radical Behaviorism & Peircean Pragmatism (Nick Thompson, Eric P. Charles)

- Hard Computational Physicalism & State Machines (Marcus Daniels, Glen Ropella)
- Non-Equilibrium Thermodynamics and Theoretical Physics (Eric Smith)
- Anthropology, Phenomenology & Hermeneutics (Dave West)
- Experimental & Scale-Appropriate Physics (Bruce Sherwood)
- Mathematics, Category Theory, abstract algebra, topology (Jon Zingale)
- Cognitive Science, Neuroscience, and Applied Mathematics/Statistics (Frank Wimberly)
- Complex Adaptive Systems, Ecological Modeling, High-Performance Simulation (Steve Smith)
- Computer Science (Russ Abbott)
- Agent-based Modelling (Douglas Roberts and Owen Densmore)

This is the group that served as an inspiration to start thinking about the rather philosophical content of this book. I joined the mailing list more than 20 years ago looking for others who are interested in these topics. They did not kick me out so I stayed. The group meets each Friday morning at coffee houses in Santa Fe for related news and developments. The discussions cover a wide range of topics spanning complexity science, computational modeling, evolutionary psychology, cognitive science, and philosophy of mind. Since the Covid pandemic when people started to work remotely there have also been regular virtual meetings. Some members of the mailing list participate remotely like me, but most are locals in Santa Fe.

Because some members have ties to the interdisciplinary Santa Fe Institute (SFI) and local tech and research communities, discussions regularly bridge technical mechanics with high-level philosophical inquiry. I have been a FRIAM list member for more 20 years now. Since I cannot remember everything we have discussed, and I was not part of all discussions, I used AI in form of ChatGPT and Gemini to make an analysis of the archive. According to the AI analysis the topics we discussed included

Complexity Science & Systems Theory

- **Emergence & Downward Causation**: How macro-level rules, structures, or properties emerge from micro-level interactions—and whether higher-level structures can exert "downward causation" back onto their constituent parts.

- **Constraint Hierarchies & Degrees of Freedom**: How systems (biological, social, or mechanical) restrict or expand their available operational trajectories.
- **Path Dependence & Self-Organization**: How non-reversible historical trajectories shape the evolution of complex adaptive systems, non-equilibrium thermodynamics, and biological systems.
- **Agent-Based Modeling (ABM) & Cellular Automata**: Using computational simulations to model complex collective behavior, artificial life, and dynamic feedback loops.
- **Scaling Laws & Network Theory**:
  Applying universal mathematical regularities, scale-free networks, and information theory to physical, biological, and social systems.

Artificial Life, Artificial Intelligence & Collective Intelligence

- **Machine Agency & AI Consciousness**:
  Comparing modern machine learning models (e.g., neural networks, LLMs) to biological cognition, exploring whether artificial agents exhibit genuine agency or internal self-models.
- **Swarm Intelligence & Distributed Control**: How decentralized groups of simple agents coordinate without central control, forming emergent social structures, laws, or consensus.
- **Computational Irreducibility**: Stephen Wolfram's concept that certain deterministic computations cannot be “shortcutted”, requiring step-by-step execution to determine the outcome—often applied in discussions about agency and decision-making.

Psychology & Behavioral Science

- **Radical & Ecological Behaviorism**: Evaluating mind and agency through direct, observable interaction with the environment (drawing on C.S. Peirce, J.J. Gibson, and E.B. Holt) rather than treating mind as an invisible internal state.
- **Self-vs.-Other Discrimination**: The evolutionary origins of agency detection—why organisms evolved mechanisms to distinguish actions initiated by themselves versus external entities.

- **Locus of Control & Trait Agency**: Operationalizing human agency as a continuous behavioral spectrum ranging from reactive passengers to proactive, adaptive decision-makers.
- **Evolutionary Psychology & Teleology**: Exploring whether mental concepts, intentionality, and moral frameworks evolved as functional adaptations, exaptations, or spandrels.

Philosophy

- **Free Will vs. Determinism**: Debating whether agency is compatible with physical law, defined through degrees of freedom, computational irreducibility, or evolutionary necessity.
- **The Hard Problem of Consciousness & Qualia**: Examining subjective experience through the lenses of self-referential computation, path-dependent emotional state histories, or category errors in language.
- **Pragmatism & Hermeneutics**: Utilizing C.S. Peirce's pragmatism, phenomenology, and object-oriented philosophy to evaluate meaning, lived narrative, and human experience against pure physical reductionism.
- **Language Hazards & Category Errors**: Analyzing how the structure of human language trickles into science, often creating false dichotomies (e.g., mind vs. body, internal vs. external).

Being a member of this fascinating discussion group was interesting and sometimes amusing. For example Doug Roberts who wrote in 20009 in one post:

> "I think I get it. Layers of abstraction, piled upon more layers. That's the answer. I forget what the question was. I can get behind that (hey, I work for a living too) [...] What I would really love to see is a concise statement of what the added value that this proposed approach is supposed to bring to the table."

Or Glen Ropella, who wrote in a 2025 discussion about "extended mind vs vibe writing" wrote "WTF are we even doing?", abbreviated it WTFAWD and proposed that the answer is basically "the same thing biology's been doing the whole time — wiggling around trying not to decohere". And then he continues to describe how he uses LLMs like this: "My own use of LLMs is almost entirely 'Say it again ... Say it again in a different way ... Again ... Again! ... Again, damnit!'"

Another example is during intense debates on consciousness or free will, whenever a discussion threatens to stall out on unexplainable physics, someone in the FRIAM inevitably invoked the “Quantum Woo” alert. When the group is spiraling into defining free will as “something that happens where there are no reasons why it happened” usually someone steps in to stop the debate from diving into “quantum woo”, warning that using quantum randomness to explain human choice never seemed to get us anywhere. In the FRIAM group we try to avoid “quantum woo”.

The two interaction patterns that emerged are

FRIAM’s 1st law:
We take difficult philosophical problems and worry them to death until someone produces either a profound insight, a completely insane hypothesis, or a phrase like ‘quantum woo.’”

FRIAM’s 2nd law:
“If a theory is suspiciously elegant, somebody on the list will throw a shoe at it.”

The members

A single “thank you” in the preface or the acknowledgements would not be enough to give credit for all the interesting ideas I stumbled upon in this group over time. We have discussed nearly all the topics of this book. Although the solutions I try to sketch in the following chapters are mostly my own thoughts and views developed later on my own after reading additional books – I am not sure if there is any idea that really belongs only to one person, though – they are in one way or the other based on all our discussions we had in the last two decades. Without all these discussions I probably would not have started to think about them. This does not mean the answers in this book are perfect, quite the contrary, but maybe they in turn can help to inspire others to find even better answers to the big questions.

Thus, I will try to give you a small overview of the prominent and active list members of the FRIAM (The Friday Morning Applied Complexity Coffee) Group in Santa Fe, Nex Mexico. It is an unusual way to give credit, but if this book should contain interesting ideas, and I hope it does, it would be unfair not to do it.

Of course, this matters only for the ideas that turn out to be relevant and true. Isaac Newton wrote many things about alchemy, the precursor of chemistry, but he did it in secret and kept it hidden from his contemporaries and colleagues. Today this work has been forgotten, and he is known as the founder of classical mechanics in physics, for inventing calculus and for his laws of motion.

He wrote the famous phrase "If I have seen further, it is by standing on the shoulders of giants" in a letter to his rival and colleague Robert Hooke, who is today in physics known for Hooke's law. Today we think of Newton’s work on calculus and the laws of motion if we hear this, but since he wrote even more on alchemy than on physics (allegedly about a million words on alchemy over his lifetime), did he think of his own contributions when he wrote the letter in 1675 in the same way as we do now?

Therefore, please forget the bad alchemy parts and remember the good scientific parts if there are any. If I accidentally should have looked at one place or another further than others in the book that follows, it is because I have been standing on the shoulders of these people who I listened to in the last two decades (plus of course the countless authors of all the books in the references):

Stephen Guerin, founder. Stephen Guerin is central to both the infrastructure and intellectual ecology of FRIAM. He founded RedfishGroup in 1991 and became involved in cognitive science, distributed systems, agent-based modeling, visualization, and complex adaptive systems. He worked with BiosGroup and participated in Stuart Kauffman’s research group. Stephen manages the FRIAM mailing list. More importantly, his role is that of organizer and intellectual catalyst: he repeatedly brings interesting phenomena, papers, simulations, and questions into the group and lets the collective discussion take them somewhere unexpected. A representative pattern is his posting of material about slime molds and collective behavior, which leads into questions about self-organization and group selection. Another humorous example concerns the phrase “batshit crazy,” which he investigated after a discussion involving Glen’s use of “wingnut.” Stephen’s role makes FRIAM itself look like a complexity experiment: provide infrastructure, connect diverse agents, avoid imposing a central intellectual agenda, and see what patterns emerge.)
→ his topics include complexity science, agent-based modeling, visualization, self-organization, Santa Fe community-building.

Owen Densmore, founding member. Owen Densmore is a veteran software engineer whose career included Xerox, Apple, and Sun Microsystems. His work included the Macintosh and Lisa era, network multimedia, Java, PostScript, VLSI, RFID, embedded Java, and later agent-based modeling and software development associated with Redfish. At Apple he worked on WYSIWIG printing and was on stage when Steve Jobs presented the Macintosh to Apple shareholders, together with Andy Hertzfeld, Randy Wigginton, Bill Atkinson, Bruce Horn, Burrell Smith, and Rony Sebok. Owen therefore provides a direct link between early personal computing and FRIAM's later interest in artificial life, simulation, and complex systems. His software models became practical tools for exploring collective behavior and emergence, complementing the more theoretical contributions of other members)
→ software architecture, software engineering, agent-based modeling, simulation, computation, real-world engineering.

Douglas (Doug) Roberts (1951-2018) Douglas grew up in Los Alamos, New Mexico. He has worked as a computer scientist at several national laboratories including Los Alamos National Laboratory and RTI International. He played the saxophone and used to ride his BMW motorcycle to places like Alaska. His email identity "parrot-farm.net" was not merely an eccentric label: he genuinely maintained a large parrot household. He and his wife took care of the 15 parrots, 40 peacocks, and an ever-changing population of stray cats that call the 200-year-old adobe house they live in home. He was devoted to the welfare of his birds and never could resist taking in a new rescue parrot. A 1992 account from the bird community describes seven parrots—Sancho, Jezabel, Maynard, Igor, Spike, Maggie, and Bandit—and Douglas carefully observes their interactions before going to work at Los Alamos. He concludes, "No wonder our neighbors think we're nuts...". This is almost a miniature of the FRIAM ethos: take an ordinary system, observe it closely, notice emergent behavior, and report it with humor. He also was a world-class computer programmer, writer, and blogger, and joined FRIAM in 2005
→ his interests included software architecture, software engineering, agent-based modeling, and simulation

Nicholas (Nick) Thompson
Inquisitive, persistent, intellectual, and playfully contrarian. Nick often acts as the list's primary conversational catalyst. He possesses a classic academic warmth paired with an unyielding commitment to Charles S. Peirce's pragmatism. He loves throwing intellectual hand grenades into discussions — challenging long-held

assumptions about the "mind" or "inner life" — and delights in forcing list members to defend their definitions. Rather than settling disputes, he thrives on keeping foundational inquiries alive.

- Background: Professor Emeritus of psychology Clark University.
- Intellectual Stance: C.S. Peirce's pragmatism, naturalistic behaviorism, and evolutionary theory.
- Core Themes on FRIAM: Frequently initiates foundational discussions on agency, intent, and consciousness. Pushes the group to view mental concepts as evolved adaptive behaviors rather than hidden internal states.
- Interests: ethology, psychology, philosophy, intentionality, evolution.

Eric P. Charles
Direct, candid, highly analytical, and refreshingly pragmatic. Eric brings an energetic, no-nonsense rigor to the list. He has a keen eye for cutting through fluff, jargon, and "armchair philosophy." His communication style is sharp, witty, and grounded; he acts as the list's empirical conscience, constantly pulling high-flying abstract theories back down to observable behavior and concrete operational definitions.

- Background: Experimental psychologist, personnel psychologist (Department of Justice), and adjunct instructor at American University.
- Intellectual Stance: Radical behaviorism and ecological psychology (J.J. Gibson, E.B. Holt).
- Core Themes on FRIAM: Serves as the list's operational skeptic and meeting chronicler. Consistently advocates for "sciencing the problem" and tethering complex philosophical concepts back to empirical behavioral testing (e.g., "rats in mazes").
- Interests: psychology, philosophy, perception, cognition.

Roger Critchlow Roger is another important early participant. He is curious, skeptical, dryly humorous, and relentlessly empirical. Roger is one of FRIAM's intellectual scouts: he brings in new papers, discoveries, technologies, and oddities, but rarely accepts their claims at face value. He has a talent for asking the deceptively simple question that exposes what a theory has left unexplained. His characteristic role is that of a curator or intellectual scout: he continually finds papers, news stories, technical developments, and oddities and throws them into the FRIAM ecosystem. His interests range across protein folding, AI, complexity, climate, information, brain mapping, governance, mathematics, and technology. A representative post about AI applied to protein sequences led the group toward protein structure, AlphaFold, and biological organization. His humor tends to be dry and one-line. A memorable example is his 2023 question about AI-generated digital replicas: “Have you added a ‘do not bot me’ clause to your will?”. Roger’s function is important historically: he acts almost like a human RSS feed, constantly supplying the group with new objects to think about

- Background: Biologist and computational/complexity researcher; early work included food-web theory and biological systems.
- Intellectual Stance: Empirical, mechanistic, skeptical of hype and vague explanations; interested in networks, emergence, and the mechanisms underlying complex patterns.
- Core Themes on FRIAM: AI and scientific discovery, protein folding, biological organization, networks, complexity, brain organization, information, climate, and the question of what mechanisms generate observed complexity.

Robert J. Cordingley Robert represents engineering, industry, management, and public affairs. He is pragmatic, experienced, systems-oriented, and socially conscious. He brings the perspective of someone who has spent a career dealing with complicated real-world systems rather than purely abstract models. He tends to connect technical ideas with their consequences for institutions, society, and human behavior, while retaining a playful interest in philosophical puzzles.
The early FRIAM description identifies him as a manager/president. Later material describes him as a retired chemical engineering fellow after 33 years in the chemical industry, and as president of 350 Santa Fe. His FRIAM interventions often connect technical and philosophical questions to real-world institutions. A memorable 2021 discussion invokes Henry II’s “Will no one rid me of this turbulent priest?” to explore indirect incitement, speech acts, and responsibility. That example is relevant to intentionality: an utterance need not explicitly command an action to function as an intentional cause within a social system)

- Background: Chemical engineer and longtime industrial practitioner; retired chemical-engineering fellow after a long career in the chemical industry; later active in Santa Fe civic affairs.
- Intellectual Stance: Pragmatic systems thinking, engineering realism, and attention to causation, responsibility, and social consequences.

- Core Themes on FRIAM: Complex social systems, technology and public policy, climate and energy, causation, responsibility, language, political behavior, and the gap between abstract ideas and their real-world consequences.

John Kennison (joined FRIAM in 2008), retired math professor, Professor Emeritus at Clark University. Quiet, analytical, mathematically rigorous, and philosophically careful, with a wonderfully understated sense of humor. John has a particular talent for noticing when apparently innocent language is smuggling philosophical assumptions into an argument. Rather than merely answering a question, he often asks whether we have formulated the question correctly in the first place.

- Background: Mathematician; PhD from Harvard; professor and researcher in category theory and topology, with interests extending into mathematical descriptions of complex systems.
- Intellectual Stance: Formal rigor combined with conceptual skepticism; particularly attentive to the relationship between mathematical structures, language, models, and phenomena.
- Core Themes on FRIAM: Mathematics and cognition, emergence, formal models, intentionality, consciousness, category theory, topology, Rosen, and whether concepts such as “goals” or “cognition” are legitimately applicable to biological systems.
- Interests: mathematics, category theory, algebra, topology

Carl Tollander (1950-2024) was a longtime Santa Fe resident. Like Richard Feynman he had a passion for drumming. He was quiet, practical, constructive, and strongly oriented toward experimentation. Carl represented the maker and systems-builder side of FRIAM: when an abstract idea becomes interesting, his instinct is often to make it concrete, build something, simulate it, or see what happens when real agents interact. His style is less philosophical than some members, but his contribution is fundamental to FRIAM's experimental character.

- Background: Engineer and applied-complexity practitioner associated with BiosGroup and agent-based/adaptive-systems work.
- Intellectual Stance: Experimental, engineering-oriented, pragmatic, and interested in understanding complex behavior through construction and simulation.

- Core Themes on FRIAM: Agent-based modeling, adaptive systems, emergence, robotics, technology, distributed behavior, simulation, and practical experimentation with complex systems.

Glen Ropella
Disciplined, meticulous, formal, and deeply probing. Glen exhibits a rigorous engineering mindset combined with a deep appreciation for formal systems and cybernetics. He is exceptionally skilled at dissecting arguments, identifying flawed metaphors, and demanding executable logic. His communication is structured, analytical, and uncompromising when it comes to clarity of terminology and state-machine transitions.

- Background: Computer scientist, agent-based modeler, and founder/lead at Tempus Dictum.
- Intellectual Stance: Formal logic, computational theory, and complex systems modeling.
- Core Themes on FRIAM: Evaluates philosophical debates (free will, agency, self-organization) through executable models, state machines, and formal feedback loops.
- Interests: complex systems, computer science, emergence, information, biological organization.

Bruce Sherwood joined the group and the list in 2011. He is a physicist, and a Santa Fe resident. Clear, unyielding, practical, and candidly physicalist. Bruce brings the disciplined clarity of an experimental physicist to list debates. He has a knack for cutting through runaway abstract speculation by testing theories against the uncompromising realities of physics and biological determinism (e.g., "Bruce's Ice Cream" preference axiom). His communication style is direct, lucid, and grounded, acting as an empirical anchor when discussions drift toward uncaused agency or "quantum woo".

- Background: Physicist, professor emeritus, and co-creator of *VPython* (a 3D programming environment widely used in physics education).
- Intellectual Stance: Hard physicalism, computational physics, and causality.
- Core Themes on FRIAM: Acts as a grounding physicalist voice on the list, testing philosophical claims against known physical laws and highlighting deterministic biological/physical counterexamples (e.g., "Bruce's Ice Cream" preference axiom)

- Interests: computational modeling and emergent behavior

Tom Johnson (Tom Johnson is a journalist associated with the Institute for Analytic Journalism in Santa Fe. His role differs from that of the mathematicians and programmers: he repeatedly brings the messy outside world—journalism, government, public data, technology, and civic institutions—into the discussion.. "Analytic journalism" is itself compatible with FRIAM: instead of merely reporting what happened, investigate the data, the system, the evidence, and the assumptions behind the story. His humor can be dry. In one exchange he says: "Don't fact-check jokes. Take the laugh and run." This produced a playful discussion about irony, stereotypes, and journalism. Tom also acted as a civic reality check, directing abstract technological discussions toward actual Santa Fe government agencies and local institutions.)
→ interested in journalism, data, government and public reality.

Pamela McCorduck (1940-2021). Pamela McCorduck contributed less frequently than many members, but her intellectual significance was much greater than her message count suggests. She was a major historian and interpreter of artificial intelligence. Her 1979 book "Machines Who Think" is a foundational history of AI. She interviewed many of the field's founders, including Herbert Simon, Allen Newell, Seymour Papert, Hubert Dreyfus, Joseph Weizenbaum, Bruce Buchanan, and others. Pamela and her husband, computer scientist Joseph Traub, bought a house in Santa Fe in 2002 and divided their time between Santa Fe and New York. Her later books, including "The Edge of Chaos" and "Bounded Rationality," reflect the Santa Fe complexity environment. Her perspective on AI was unusually broad: she had seen the field's philosophical arguments at their birth and could connect computation to questions about human intelligence, culture, cognition, and meaning. In a later interview she described the Santa Fe Institute as giving her a vocabulary for explaining things she had already observed in AI. This makes her a natural bridge between AI history, complexity, and philosophy.)
→ AI history + humanities + philosophy + the human meaning of technology

Russell Standish (Sydney, Australia)
Russell Standish is an Australian mathematician/computer scientist and complexity researcher associated with UNSW Sydney. He represents the mathematical and computational side of FRIAM. His research includes complexity, networks, artificial life, computation, and information-based approaches to complexity. His

2008 paper "Concept and Definition of Complexity" distinguishes complexity as a quality related to our ability to understand something from complexity as a quantity that might be measured. His FRIAM style tends toward formalization: when someone says a system is "complex," the Russell move is to ask in what sense, whether it can be measured, and what the model predicts. A memorable 2006 exchange about the colors of the rainbow has him writing: "To be quite frank, I cannot distinguish a separate color between blue and violet - does that mean my cones are defective?" The discussion then moves through perception, color categorization, mathematics, and culture. Russell also wrote "Ants are not Conscious" (2008), illustrating his willingness to distinguish sophisticated behavior from consciousness rather than automatically moving from cognition to subjective experience)
→ mathematics, computation, complexity measures, artificial life.

Pieter Steenekamp (South Africa)
Intellectually ambitious, technically sophisticated, enthusiastic, and self-questioning. Pieter is comfortable moving between engineering, physics, AI, and philosophy, and he has a characteristic willingness to entertain very large ideas while simultaneously asking whether the underlying mechanism really supports them. His self-deprecating humor and readiness to admit confusion make him an unusually open interdisciplinary participant.

- Background: Engineer, programmer, entrepreneur, and founder of Rand Controls; independent consultant with interests spanning engineering and technology.
- Intellectual Stance: Technologically optimistic but intellectually skeptical; uses engineering and physical reasoning to test philosophical and scientific claims.
- Core Themes on FRIAM: AI, consciousness, intentionality, scientific discovery, fundamental physics, emergence, AI-assisted programming, technology, and the distinction between a mathematical description and the physical process it describes.

Steve Smith
Holistic, expansive, integrative, and culturally grounded. Steve possesses a wide-angle intellectual lens, effortlessly linking abstract physics and complexity models to real-world ecosystems, human organizations, and social institutions. His posting style is reflective and synthesis-driven, acting as a natural bridge-builder across

disciplines who consistently reminds the group of macro-level contexts and constraint hierarchies.

- Background: Complex systems researcher, entrepreneur, and long-time Santa Fe Institute (SFI) affiliate.
- Intellectual Stance: Dynamic systems, network theory, and structural hierarchy.
- Core Themes on FRIAM: Focuses on constraint hierarchies, downward causation, and how macro-structures (organizations, institutions, ecosystems) constrain or modify the degrees of freedom of micro-agents.

Russ Abbott

Systematic, inquisitive, intellectually ambitious, and structural. Russ approaches discussions with the structured mindset of a software architect and computer scientist. He is fascinated by the mechanics of emergence—how higher-level entities and rules arise from lower-level computational or physical substrates. His posting style is analytical and conceptual, constantly challenging the list to define precisely how abstraction layers interact and operate in complex systems.

- Background: Professor Emeritus of Computer Science at California State University, Los Angeles.
- Intellectual Stance: Computer science, emergence theory, software architecture, and complex systems.
- Core Themes on FRIAM: Focuses heavily on the nature of emergence and entity creation in software and complex systems. Russ frequently argues that computer science provides a concrete framework for understanding how higher-level level-spanning entities and rules emerge from lower-level physical/computational substrates.

David (Dave) West

Expressive, reflective, iconoclastic, and deeply empathetic. Dave serves as the essential humanistic balance to the list's hard physicalists and computationalists. He writes with a rich, narrative style that values meaning, culture, and first-person lived experience. Intellectually fearless and comfortable pushing back against pure reductionism, he brings warmth, philosophical nuance, and a design-oriented perspective to technical debates.

- Background: PhD in anthropology, software design theorist, object-oriented programming pioneer, and anthropologist/philosophical thinker.

- Intellectual Stance: Hermeneutics, phenomenology, object-oriented philosophy, and design thinking.
- Core Themes on FRIAM: Counterweight to pure reductionism and physicalism. Advocates for taking first-person lived experience, context, narrative, and culture seriously as primary dimensions of living systems.

Marcus Daniels
Sharp, laconic, highly technical, and uncompromisingly logical. Marcus writes with razor-sharp efficiency and computational precision. He doesn't waste words on fluff or sentimentality, approaching philosophical debates with the cold, clear logic of a systems architect. His style is incisive and direct, consistently testing theories against the hard realities of code, physical hardware, and computational feasibility.

- Background: Computational scientist, software engineer, and simulation builder (notably involved with Swarm simulation software and high-performance computing).
- Intellectual Stance: Computational physicalism, algorithm design, and simulation theory.
- Core Themes on FRIAM: Views mental processes, consciousness, and choice through a technical computational lens—aligning with illusionist/computational models where consciousness is an internal diagnostic executing on physical hardware.

Frank Wimberly
Patient, collegial, mathematically precise, and understated. Frank writes with the calm, measured authority of a seasoned mathematician and cognitive scientist. He is exceptionally skilled at bridging abstract mathematical structures with empirical neuroscience and physics. His tone is consistently polite, balanced, and constructive; he rarely engages in fiery rhetoric, preferring to guide debates toward logical consistency, scale-appropriate physics, and clear reasoning.

- Background: Mathematician, computer scientist, and former faculty member at Carnegie Mellon University (CMU).
- Intellectual Stance: Mathematical modeling, empirical science, and systemic physicalism.
- Core Themes on FRIAM: Bridges mathematical rigor with empirical science, regularly urging the group to ground emergent macro-properties in concrete mathematical logic, neuroscience, and scale-appropriate physics.

Eric Smith
Rigorously deep, intellectually formidable, physics-driven, and foundational. Eric brings the heavy machinery of statistical mechanics and non-equilibrium thermodynamics to FRIAM discussions. His communication style is intense, dense, and deeply analytical, consistently pushing the list to abandon superficial metaphors and look at the underlying physical dynamics. He views biological complexity, life, and agency not as mysterious exceptions to nature, but as the inevitable, mathematically driven collapse of chemical and thermal gradients.

- Background: Theoretical physicist, professor at the Santa Fe Institute (SFI), and co-author of *The Origin and Nature of Life on Earth*.
- Intellectual Stance: Non-equilibrium thermodynamics, statistical mechanics, phase transitions, and origin-of-life biology.
- Core Themes on FRIAM: Frames living systems as driven non-equilibrium systems that collapse energy gradients. Grounds emergence in formal statistical mechanics and phase transitions.

Jon Zingale
Methodical, precise, deeply thoughtful, and intellectually humble. Jon communicates with the calm clarity of a formal mathematician. He excels at spotting hidden structural patterns and logical ambiguities that others miss. While his analytical depth is immense—frequently drawing on category theory or formal logic—his tone remains collaborative, patient, and pedagogical, often helping bridge the gap between abstract math and human intuition.

- Background: Mathematician, software developer, musician/sound artist, and educator with deep roots in Santa Fe's technical and creative communities.
- Intellectual Stance: Category theory, formal logic, abstract algebra, signal processing, and mathematical philosophy.
- Core Themes on FRIAM: frequently brings a rigorous mathematical lens to list debates, often using category theory, topology, and abstract algebra to formalize squishy philosophical concepts. He regularly examines how structures transform across domains, helping the list map relationships between software architectures, physical models, and cognitive processes. He also frequently serves as an intellectual translator on the list, breaking

down dense mathematical frameworks into approachable concepts for list members coming from biology, psychology, or sociology backgrounds.

This list was again AI generated using ChatGPT and Gemini but as far as I can see it seems to be accurate. As you can see the members include many retired professors from multiple fields, tech proficient specialists, software developers and experienced engineers - people interested in complexity science, philosophy, psychology, collective intelligence and complex systems in general.

In the beginning twenty years ago agent-based models where a hot topic. Doug Roberts and Owen Densmore for example worked on agent-based models and built their own frameworks. Today AI is of course the big topic, and AI agents are the new big thing. Humans tend to see the world as a collection of interacting agents.

The FRIAM cast in one sentence:

- **Steve Smith:** *"InerESTING theory—but let's see if it survives contact with the evidence."*
  Empirical, technically minded, skeptical, and occasionally delightfully blunt. Steve represents the engineering/physical-science instinct to test elegant explanations against reality rather than accepting them because they sound plausible.
- **Marcus Daniels:** *"That's much too neat; reality is messier than that."*
  Sharp, iconoclastic, skeptical, and fond of puncturing intellectual pretension. Marcus is one of the list's great "shoe throwers," particularly suspicious of theories that explain too much too elegantly.
- **Nicholas Thompson:** *"What does this tell us about the organism actually doing the thinking?"*
  Philosophically adventurous, empirically grounded, and endlessly curious about animal behavior and mind. Nick is FRIAM's bridge between ethology, evolutionary biology, intentionality, consciousness, and philosophy of mind.
- **Owen Densmore:** *"Let's stop theorizing and build the damn thing."*
  Practical, inventive, technically fearless, and deeply experienced in software and computing. Owen brings the hacker/engineer instinct: ideas become interesting when they can be implemented, simulated, or made to behave.
- **David West:** *"You're leaving out the human being who actually experiences all this."*
  Expressive, reflective, iconoclastic, and deeply empathetic. Dave provides the essential humanistic counterweight to reductionism, emphasizing lived experience, meaning, culture, narrative, design, and interpretation.
- **Frank Wimberly:** *"There's probably another way of looking at this."*
  Reflective, intellectually playful, independent-minded, and comfortable moving between technical and philosophical perspectives. Frank contributes a willingness to question assumptions and explore the conceptual space around a problem rather than rushing toward a single answer.
- **Glen Ropella:** *"Perhaps the interesting question is what kind of system could produce this behavior."*
  Analytical, inventive, mathematically inclined, and fascinated by emergence. Glen repeatedly turns observations about cognition, mathematics, biology, and complex systems into questions about how the underlying machinery could generate them.

- **Tom Johnson:** *"That's interesting—but what does the data actually say?"*
Curious, observant, civic-minded, and journalistically skeptical. Tom brings the outside world into FRIAM, connecting complexity and technology to journalism, government, public data, and actual social institutions.
- **Jochen Fromm:** *"Maybe the apparent complexity is hiding a surprisingly simple computational principle."*
Systematic, technically sophisticated, and fascinated by the relationship between computation and emergence. Jochen brings a strong computational perspective to questions about life, consciousness, complexity, and self-organization.
- **Eric P. Charles:** *"Before we accept that explanation, let's look at what the system is actually doing."*
Analytical, skeptical, technically minded, and willing to challenge assumptions. Eric's contributions tend toward careful examination of mechanisms rather than accepting philosophical interpretations at face value.
- **Roger Critchlow:** *"Interesting—but what mechanism actually produces it?"*
Curious, skeptical, dryly humorous, and relentlessly empirical. Roger is FRIAM's intellectual scout, continually bringing new discoveries and ideas to the group while asking what mechanism lies underneath the phenomenon.
- **Douglas Roberts:** *"I think I get it. Layers of abstraction, piled upon more layers. That's the answer. I forget what the question was"*
Brilliant, blunt, independent, technically formidable, and gloriously intolerant of nonsense. Doug represents the old-school Los Alamos programmer: if an argument is wrong, inefficient, or annoying, he may not bother constructing a diplomatic rebuttal—he'll throw a shoe at it.
- **Jon Zingale:** *"Let's work out what the system actually permits before we speculate about what it means."*
Analytical, technically rigorous, and oriented toward physical and computational mechanisms. Jon brings a scientific instinct for constraining philosophical speculation with models and concrete system behavior.
- **Gillian Densmore:** *"But what does this look like when people actually have to live with it?"*
Observant, practical, empathetic, and grounded in human consequences. Gillian provides an important social and human perspective alongside Owen's strongly technical orientation.

- **Stephen Guerin:** *"Here's something interesting—let's throw it into the group and see what happens."*
  Curious, entrepreneurial, playful, and extraordinarily good at connecting people and ideas. Stephen is the gardener of FRIAM: he provides the infrastructure, introduces provocative material, and lets the collective intelligence self-organize.
- **Pieter Steenekamp:** *"This is fascinating—but is it really doing what we think it is doing?"*
  Technically ambitious, enthusiastic, self-questioning, and comfortable moving between engineering, physics, AI, and philosophy. Pieter combines technological optimism with a strong instinct to test grand claims against mechanism.
- **Merle Lefkoff:** *"Complex systems are not just machines; people and institutions are part of the system."*
  Systems-oriented, socially conscious, interdisciplinary, and interested in applying complexity thinking beyond the laboratory. Merle brings human conflict, cooperation, governance, and social transformation into the complexity conversation.
- **Russell Standish:** *"Can we define that precisely enough to measure it?"*
  Mathematical, skeptical, intellectually disciplined, and quietly playful. Russell brings the formalist instinct to FRIAM: define the concept, distinguish quantity from quality, and ask whether the proposed complexity or consciousness can actually be characterized.
- **Phil Henshaw:** *"What if the thing we're calling a problem is actually a property of the larger system?"*
  Systems-oriented, ecological, energetic, and willing to challenge conventional boundaries. Phil tends to look beyond individual components toward flows, networks, environmental constraints, and the larger structures in which phenomena occur.
- **Robert Holmes:** *"Let's not forget that complicated systems have history."*
  Reflective, technically curious, and interested in the relationship between structure, dynamics, and historical development. Robert contributes to FRIAM's broader interest in how complex systems acquire the properties they have.
- **Barry MacKichan:** *"Let's make the abstract idea concrete enough to argue about."*
  Practical, technically minded, and constructive. Barry's contribution fits the FRIAM engineering tradition: turn vague concepts into systems, mechanisms, examples, or implementations that can actually be examined.

- **Russ Abbott:** *“Maybe the distinction we're arguing about is one our model invented.”*
  Philosophically sophisticated, computationally oriented, and interested in the relationship between formal systems and philosophical concepts. Russ brings computer science into questions of mind, agency, abstraction, and emergence.
- **Carl Tollander:** *“Can we build it and see what happens?”*
  Quiet, practical, constructive, and experimentally minded. Carl represents FRIAM's applied-complexity tradition: agent-based models, adaptive systems, technology, simulation, and the belief that building a system can reveal things that abstract reasoning misses.
- **Pamela McCorduck:** *“Before deciding what AI means, remember that humans have been arguing about this question since the beginning of AI.”*
  Historically informed, intellectually broad, humane, and deeply interested in the meaning of artificial intelligence. Pamela brings the extraordinary perspective of someone who personally knew and interviewed many of AI's founders.
- **Eric Smith:** *“Don't start with the answer; start with the process that could have produced it.”*
  Deeply theoretical, scientifically ambitious, and fascinated by the origins of complexity. Eric brings thermodynamics, statistical physics, information, chemistry, evolution, and the origin of life together in an attempt to understand how organized complexity becomes possible.
- **Gary Schiltz:** *“There's probably a practical engineering problem hiding inside this philosophical question.”*
  Technically minded, pragmatic, and solution-oriented. Gary contributes the engineering instinct to translate broad ideas about systems and complexity into mechanisms that could actually operate.
- **John Kennison:** *“Before answering that, are we sure the concepts we're using are appropriate to the phenomenon?”*
  Quiet, mathematically rigorous, philosophically careful, and wonderfully understated. John is particularly good at detecting when words such as *goal*, *cognition*, or *consciousness* are importing assumptions that have not been justified.
- **Bruce Sherwood:** *“Let's simulate it and find out.”*
  Computational, educational, pragmatic, and strongly committed to making abstract concepts experimentally accessible. Bruce represents the tradition of using computation not merely to calculate answers but to make complex phenomena observable and manipulable.
- **Sarbajit Roy:** *“There's a mathematical structure underneath this somewhere.”*
  Analytical, technically rigorous, and interested in finding formal structure

beneath apparently messy phenomena. Sarbajit brings mathematical abstraction into the interdisciplinary conversation, providing another route from complexity toward something that can be precisely reasoned about.

- **Roger Frye:** *“The interesting part is what happens when the pieces interact.”*
  Systems-minded, curious, and interested in interaction and emergence. Roger's FRIAM perspective fits the group's broader instinct to look beyond isolated components toward the behavior produced by relationships among them.

Together we tried to find solutions for the big philosophical questions. Each considered the problem from his own perspective. The psychologists Nick and Eric explained us the insights of the founders of psychology. If you find William James, Charles Peirce or John Dewey mentioned in this book, then the roots can be found in Nick’s emphasis of Peirce and his contemporary philosophers.

FRIAM repeatedly returned to the questions of top-down causation and bottom-up emergence, the problem of free will, subjective consciousness, the hard problem of consciousness, the mind-body problem, intentionality, representation, and the status of models.

- When we describe something successfully by a model, have we explained it — or merely constructed a useful representation?
- Can models become dangerous if we mistake them for reality?
- Emergence. Can a higher-level phenomenon be completely explained by its components?
- Free-Will. What kinds of physical systems can control their own future behavior?
- Intentionality. How can a completely physical system be about something?
- Can higher-level phenomena be completely explained by their components, or do organization, history, feedback, and context matter?
- How can physics underlie the mind? How does meaningless physical causation become meaningful directed activity?
- Is mathematics discovered or invented?
- What is consciousness?

No questions were too big for us, and we tried to tackle them without using mysterious entities or obscure theories. Like Daniel Dennett, many FRIAM participants were willing to tackle apparently impossible philosophical questions using ordinary language, empirical examples, evolutionary reasoning, computation, and careful analysis of what people do.

The picture that emerged was that some complex systems cannot be reduced to the components without looking at the history. Systems generate genuinely new behavior through continuous interaction, feedback, history and context in open-ended systems, which is hard to explain by simple closed-form explanations. Cervantes said in his book “Don Quixote" that "there is no book so bad... that it does not have something good in it”. In the group the people are so curious that there is no thing so ordinary that it is not worth thinking about it. The advantage is that if nothing is too ordinary to investigate, then you can discover the answer to the big questions if they are hiding in plain sight behind ordinary things. And this turns out to be surprisingly true. The ordinary things are often the fundamental things that have been around for such a long time that they are taken for granted. And these things are the places where the answers to the big questions are hiding in plain sight, and if we look closely enough, we can spot them. The most ordinary things are often the things where the answers to the big questions are hiding in plain sight.

On Intentionality

Nicholas Thompson's paper "Intentionality is the Mark of the Vital" drew my attention to the mind-body problem and to philosophical terms such as intentionality. Intentionality in philosophy has not the same meaning as in everyday language. In philosophy it means the philosophical concept of "aboutness," or how mental states refer to or represent things in the world.

In the group we have approached intentionality through a mix of evolutionary biology, behavioral psychology, pragmatism, and complex systems modeling. Rather than viewing intentionality as a spooky "internal mental force" hiding in the brain, list members generally debate it as a descriptive pattern of goal-directed behavior, an ecological adaptation, or a function of computational feedback systems.

The Naturalistic & Behavioral View: "Descriptive, Not Explanatory"

Key Figures: Nick Thompson, Eric P. Charles

- **Intentionality is Immanent in Biology:** Nick Thompson and Eric Charles argue that intentionality is not a mysterious capacity unique to human consciousness. Instead, concepts exhibiting the logical properties of intentionality are commonplace across biology. Even simple organisms exhibit directedness toward their environment. Intentionality is not an invisible internal cause; it is a descriptive label for observable, high-order, goal-directed patterns of behavior over time that match an organism to its environment.
- **Rejection of Hidden Mental Causes:** Drawing on C.S. Peirce's pragmatism and naturalistic behaviorism, Nick and Eric argue that "intentions" should not be invoked as invisible internal causes of behavior.
- **Macro-Patterns of Behavior:** Intentionality is treated descriptively, not explanatorily. When we say an agent "intended" to do something, we are describing a high-order, observable pattern of behavior over time that matches the organism to its environmental circumstances.

The Computational & Cybernetic View: Goal-Directed State Machines

Key Figures: Glen Ropella, Marcus Daniels

- **Feedback Loops & Teleonomy:** From a computational perspective, intentionality is framed through feedback loops, objective functions, and target states in executable models.
- **Internal Models:** An artificial agent or biological system exhibits intentionality when it maintains an internal model of its environment to evaluate and adjust its trajectories toward a goal or target state. Intentionality is thus the operational feedback loop between an internal model and external state transitions.

The Linguistic / Semantic Distinction: Intentionality vs. Intensionality

Key Figures: Jon Zingale, Nick Thompson
List discussions frequently clarify the logical and linguistic boundaries of the term:

- **Intentionality (with a 't'):** The philosophical problem of directedness—how a state or sign refers to an object or goal.
- **Intensionality (with an 's'):** The logic of meaning, context, and definition (intension vs. extension) in formal language.

Members often debate whether the "aboutness" of mental states is simply a property of how symbolic systems and languages handle reference across contexts.

The Complex Systems & Ecological View: Affordances and Constraints

Key Figures: Steve Smith, Dave West

- **Landscapes of Affordances:** Rather than treating intentionality as an isolated internal state inside a head, ecological and systems thinkers on the list view it as an interaction between an agent and its "niche".
- **Systemic Directionality:** Intentionality emerges when an agent navigates a structured landscape of constraints and opportunities (affordances), where the system's structure guides choice and direction.

On the mind-body problem

On the FRIAM mailing list, the **mind-body problem** - how mental phenomena (thoughts, feelings, intentions) relate to physical processes (brains, neurons, physiological states) - is approached with widespread skepticism toward traditional Cartesian substance dualism.

Because list members are largely complexity scientists, physicalists, behaviorists, and software theorists, the consensus broadly rejects the idea that "mind" and "body" are two separate substances. Instead, the list redefines the problem through embodied cognition, multi-level systems constraints, functional self-modeling, and radical behaviorism.

The Monist / Embodied View: "Mind is Body, Beyond the Brain"

Key Figures: Marcus Daniels, Glen Ropella, Eric Charles

- **Extended Biology:** List members frequently push back against the naive reduction of "body" to just "the brain". In discussions on biology, members point out that physiological systems like the enteric nervous system (the "gut brain"), the vagus nerve, endocrine signaling, and tactile neurons do significant computational "thinking".
- **No Ghost in the Machine:** Mind is not an intangible ghost residing inside a physical shell. Under this view, "mind" is simply the aggregate, dynamic functional output of physical biological tissue interacting with its environment.

The Pragmatist & Behaviorist View: Dissolving the Pseudo-Divide

Key Figures: Nick Thompson, Eric P. Charles

- **A Linguistic Trap:** Drawing on C.S. Peirce and Gilbert Ryle, Nick Thompson and Eric Charles treat the "mind-body problem" as a classic category error and a trap created by mentalistic language.
- **Mind as an Adverb, Not a Noun:** They argue that "mind" is not a physical object *or* a non-physical substance; it is a way of describing how an organism acts over time. To say a body has a "mind" is to describe its coherent, goal-directed, and context-sensitive behavior in an environment. Once you study the organism's observable relationship with its environment, the "mind-body problem" dissolves.

The Complex Systems & Emergence View: Level-Spanning Hierarchies

Key Figures: Steve Smith, Russ Abbott, Bruce Sherwood

- **Macro-Properties and Constraints:** Systems theorists on the list view "mind" as an emergent macroscopic property of lower-level physical interactions.
- **Downward Constraints:** Rather than treating mind and body as separate, they frame them as different levels of hierarchy within a complex adaptive system. The physical body provides the substrate (micro-level degrees of freedom), while the "mental" (macro-level organizational patterns, social rules, and habits) constrains how lower-level physical states evolve over time.

The Computational / Functionalist View: Hardware vs. Software

Key Figures: Marcus Daniels, Glen Ropella, Russ Abbott

- **Substrate Independence:** Computer scientists on the list frequently frame the mind-body problem using the hardware/software analogy. "Body" is the physical hardware/state-machine, while "mind" is the executing software algorithm or computational process.
- **Informational States:** They argue that while mental states require a physical substrate to execute, the logical structure of the software (information processing, diagnostic self-monitoring) is what defines "mind".

The Phenomenological & Hermeneutic Counterweight

Key Figures: Dave West

- **Critique of Pure Reductionism:** Dave West frequently cautions the group against reducing human experience entirely to mechanics, circuitry, or behavioral outputs.
- **Lived Experience:** He advocates for taking the first-person, lived experience of having a body seriously. Under this lens, reducing mind entirely to brain tissue or lines of code strips away essential human dimensions like meaning, narrative, and culture.

On Free Will

Two problems we often discussed are the problem of free will and the hard problem of consciousness, simply because they are the most difficult and the most interesting problems in philosophy. They are also a central part of this book. While our FRIAM discussions on free will span many threads over two decades, the core points of view expressed by various members generally group into several key perspectives:

The Complex Systems View

- **Core Idea:** Free will is an emergent property of complex, multi-scale hierarchical systems.
- **Key Arguments:**
    - Determinism at lower levels (molecules, neurons) does not dictate or fully explain behavior at higher levels (consciousness, agency)..
    - In non-linear dynamical systems, small changes lead to unpredictable macro-behavior (sensitivity to initial conditions).
    - Freedom is framed as **degrees of freedom** within a constrained state space. A human has vastly more internal states and trajectory choices than a simple organism or rule-based machine.

The Pragmatic View

- **Core Idea:** Whether free will exists is less important than how agency functions in practice.
- **Key Arguments:**
    - Free will can be defined operationally as the capacity of a system to evaluate options, project future scenarios, and alter its course of action based on internal models.
    - List Members often draw on Pierce's pragmatism or cybernetics: if acting as if we have agency is necessary for social coordination, moral responsibility, and goal-directed behavior, then it exists in a functional sense.

The Computational / Algorithmic Perspective

- **Core Idea:** Free will is tied to computational irreducibility and decision-making algorithms
- **Key Arguments:**
    - According to Stephen Wolfram's idea of computational irreducibility, even if a process is deterministic, there is no shortcut to predict its outcome other than running the computation. Thus, an agent cannot know its own future decision until it makes it, producing the lived experience of choice.
    - Discussions frequently compare human choice to artificial intelligence, exploring whether reinforcement learning agents or neural networks exhibit a form of "free will."

The Skeptical / Deterministic View

- **Core Idea:** True "uncaused" free will is an illusion incompatible with physical laws
- **Key Arguments:**
    - The universe is either deterministic (governed by physical laws) or fundamentally random at quantum scales—neither allows for a traditional "libertarian" free will.
    - Subjective feelings of conscious agency are often post-hoc rationalizations generated by the brain *after* neural processes have already initiated an action (citing Libet-style experiments and neuroscience).

The Semantic / Linguistic Frame

- **Core Idea:** The debate over free will is largely a category mistake or semantic confusion
- **Key Arguments:**
    - Members frequently point out that the term "free will" is poorly defined. "Free from what?"—free from physical causation (impossible), or free from external coercion (real and measurable)?
    - Once you carefully specify what you mean by "freedom" and "will," much of the philosophical paradox dissolves into questions about cognitive architecture and social contracts

Rather than taking a single side, in the FRIAM group we have generally tried to treat free will not as an "all-or-nothing" metaphysical switch, but as a spectrum of agency and control rooted in cognitive complexity, evolutionary feedback loops, and dynamic interaction with the environment. Free will and agency are seen in general as a continuum or spectrum of constrained "degrees of freedom" rather than a binary yes/no property. Some people have fewer cognitive degrees of freedom than others, and there are moments where we show full free will, and others where we lack it.

The FRIAM group members – Nick Thompson, Eric P. Charles, Steve Smith, Glen Ropella and Marcus Daniels have mentioned this – generally reject the traditional metaphysical framing of free will as an "uncaused, supernatural intervention" or an "all-or-nothing switch". Instead, the list consensus views agency as a **spectrum of constrained degrees of freedom** within complex, multi-scale dynamic systems. Agency increases or decreases depending on an organism's internal predictive models, cognitive complexity, and structural constraints.

A spectrum model defined by three parameters seems to be reasonable:
internal capacity, degrees of freedom and external constraints

1. **Internal Capacity:**
   How many internal states and predictive models can the system evaluate?
2. **Degrees of Freedom:**
   How many real, physical paths are available to the agent?
3. **Constraints:**
   How tightly is the agent bound by lower-level physics, biological hardwiring, or higher-level social/institutional systems? For example, when someone vows obedience to an organization or enters a higher-order social structure, they willingly subsume their agency to the institution.

On the high-end of the agency & free-will spectrum we have intelligent proactive agents who evaluate options, adapt dynamically to obstacles, and shape their environment. On the low-end of the agency & free-will spectrum there are reactive agents who behave like passive passengers governed primarily by external forces, luck, or rigid habits.

The view that emerged is that agency forms a spectrum that varies strongly on the proactivity/reactivity scale. It also scales up from basic physical particles to complex biological organisms. Free will and agency are best described through degrees of freedom and how choices and possibilities are constrained at different levels of organization. Having "more" or "less" free will can depend on the system's constraints, i.e. on the rules, habits an entity has developed and on the social and physical world it is embedded in.

As Dave West wrote in the group in August 2024 a quark or amino acid has severely restricted choices governed almost entirely by basic physical laws; its "odds" of having free will or making an alternative choice are nearly zero. If biological organization increases in complexity, the apparent range of choices expands significantly by combinatorial explosion.

The twist is that even at the higher end of the spectrum which has the largest number of choices, an agent's choices remain deeply constrained by past cumulative choices and preferences built in over time (e.g., comparing innate taste preferences vs. deliberate long-term behavioral changes).

Key Figures & Specific Positions on Free Will

Nicholas S. (Nick) Thompson (Evolutionary / Pragmatist Lens):
View: Frames free will as an **evolutionary adaptation** rather than a spandrel or illusion.
Argument: Humans evolved agency-detection mechanisms and self-vs-other discrimination because recognizing who or what initiated an action offered significant survival advantages.

Eric P. Charles (Behaviorist / Empirical Lens):
View: Demands an **empirical and operational definition** of agency over "armchair philosophizing."
Argument: Defines free will descriptively via *degrees of freedom* and system constraints (e.g., the "joining the clergy" metaphor, where an individual enters a state of subsumed agency). Argues that if free will means "completely uncaused action," it becomes synonymous with magic and loses scientific utility.

Bruce Sherwood (Physicalist / Deterministic Lens):
View: Pushes a **causal and physicalist critique** of uncaused choice.
Argument: Demonstrates that whenever an agent appears to choose freely (e.g., choosing chocolate over vanilla ice cream), deep-seated biological, psychological,

and physical causes drive that preference.

Steve Smith (Dynamic Systems / Structural Lens):
View: Analyzes agency through **constraint hierarchies and downward causation**.
Argument: Explores how individual degrees of freedom are altered, constrained, or expanded when an agent integrates into higher-order social, cultural, or institutional structures.

Glen Ropella & Marcus Daniels (Computational / Formal Logic Lens):
View: Approach agency through **computational irreducibility and executable models**.
Argument: Argue that even in deterministic state machines, an agent cannot compute its ultimate decision without executing the full algorithm, generating the operational experience of choice.

Summary Table: Perspectives on Free Will

Member / Contributor	Primary Intellectual Frame	Core View / Perspective
Nick Thompson	Evolutionary Psychology & Peircean Pragmatism	Evolutionary adaptation for self/other agency detection and social coordination.
Eric P. Charles	Behaviorism & Operational Psychology	Descriptive measurement of dynamic degrees of freedom and state constraints.
Bruce Sherwood	Physicalism & Physics	Skeptical of uncaused choice; highlights pre-existing physical and biological causes.

Member / Contributor	Primary Intellectual Frame	Core View / Perspective
Steve Smith	Complex Systems Theory	Hierarchy of constraints where higher-order social/institutional structures bound agency.
Glen Ropella	Computer Science & Formal Modeling	Agency as algorithmic self-correction, state transitions, and executable feedback loops.

On the Hard Problem of Consciousness

As FRIAM members we could not resist to discuss David Chalmers' "Hard Problem of Consciousness”. Some of us treated the "Hard Problem" - why physical processes give rise to subjective qualia - as a **pseudo-problem, a category error, or an artifact of self-referential computation**. Rather than accepting an unbridgeable metaphysical divide between mind and matter, we have viewed consciousness as an emergent property of high-dimensional feedback loops and path-dependent history.

My own position is that **path dependence** is the key to solving the hard problem. As you can read in later chapters, subjective experience (or qualia) emerges in my point of view from the non-reversible, accumulated path-dependent history of emotional and internal states over time, which bridges universal physical laws with individual uniqueness.

Key Figures & Specific Positions

Nick Thompson & Eric P. Charles (Pragmatism & Behaviorism):

View: Reject the notion of a private "inner theater."

Argument: Treat subjective experience as a **category error**. Consciousness is not an intangible internal substance, but an organism's goal-directed interaction with its environment over time.

Marcus Daniels & Glen Ropella (Computationalism & Self-Modeling):
View: Frame subjective experience as **internal self-representation**.
Argument: Align with Dennett's illusionism: when a complex computational architecture monitors its own internal state, the system's high-level summary of its state feels like *qualia* from the inside.

Dave West (Phenomenology & Narrative):
View: Counterweight to pure reductionism.
Argument: Advocates for taking first-person lived experience, meaning, and narrative context seriously, arguing that reducing consciousness entirely to neural wiring strips away essential systemic properties.

Bruce Sherwood & Frank Wimberly (Emergence & Multi-scale Physics):
View: Treat consciousness as a **macroscopic emergent property**.
Argument: Caution against scale errors (e.g., searching for qualia in individual neurons), noting that consciousness emerges across spatial and temporal scales just as wetness emerges from molecular configurations.

Subjective experience is the central problem of the hard problem. In our discussions for subjectivity and objectivity the following principles emerged: Steve Smith noticed the principle of objectivity, Glen Ropella discovered the paradox of subjectivity and Jon Zingale emphasized the principle of invariance.

Steve's principle of objectivity:
"what we call 'objective truth' is ultimately just a deeply shared, agreed-upon model rather than an absolute, independent reality". Objective truth is only a model of reality that humans collectively agree is accurate, rather than reality itself in some absolute, ultimate sense. It is less like an unshakeable bedrock existing independently in the universe and more like a high-resolution, universal map that all humans have mutually agreed to draw and use. In this sense objective truth remains subjective too because it is the perspective of humans that have mutually agreed to share this view.

Glen's paradox of subjectivity:
As we grow older, we become both more subjective and less subjective. More subjective because our subjective 1st person view becomes more unique, more specific and more singular. Less subjective because we gain an additional objective view - Nagel's objective 3rd person "view from nowhere". We can say our

perspective becomes simultaneously more personal and more objective, as we develop a unique first-person viewpoint alongside an increasingly detached third-person perspective. We become both more uniquely subjective and more capable of transcending our subjective perspective.

Jon's principle of invariance:
Reality is change; objects are patterns that persist through change. If a subject perceives reality fundamentally as a continuously changing stream, then an object can be understood as a pattern that is a persistent invariant in a changing process. Objects are patterns that persist through change. Cognition identifies objects as persistent patterns that remain invariant under transformations of the underlying stream.

Summary Table: Perspectives on Subjective Experience

Member / Contributor	Primary Intellectual Frame	Core View / Perspective
Nick Thompson & Eric Charles	Peircean Pragmatism / Behaviorism	Category error; consciousness is goal-directed environmental interaction, not an "inner movie."
Marcus Daniels & Glen Ropella	Computational Theory & Cybernetics	Self-monitoring feedback loops; qualia is the internal view of high-level state diagnostics.
Jochen Fromm (myself)	Path Dependence & Complexity Science	Subjective experience is the cumulative, non-reversible path dependence of internal emotional states.

Member / Contributor	Primary Intellectual Frame	Core View / Perspective
Dave West	Phenomenology & Hermeneutics	First-person experience and narrative meaning are primary dimensions that cannot be reduced to code.
Bruce Sherwood & Frank Wimberly	Emergent Physics & Neuroscience	Emergent macro-property across scales; searching for qualia at the micro level is a scale mistake.

I hope you have got an impression of the wide range of topics we have discussed in the last years. Many of them touched the biggest philosophical problems we know, and since we all have a slightly different background, the interdisciplinary discussions were often fruitful and interesting – unless they ended in “quantum woo”.

In the following chapters I will try to sketch my own view of these philosophical problems from intentionality to the mind-body problem: if there are any real solutions, and how they would look like from my point of view. They are a collection and extension of my blog posts about the same issues. The seed for each blog post was often one of my own posts in the FRIAM group, which was in turn often a reply to someone else in the group. Using additional books and references I have tried to understand these open and unsolved questions, but the original inspiration was often one of the discussions in the FRIAM group.

Fiction and the paradox of intentionality

Core points:

- Paradox: mental states (like thoughts or fears) can be "about" imaginary objects that do not exist
- How can a real mental state inside our heads be about external objects outside the head that are either far away or imaginary?

The paradox of intentionality

> “It is difficult to imagine how mere bits of matter inside the skull could refer to or be about something in the world beyond themselves”
> ~ John Searle in “Mind – A Brief Introduction” (Searle, 2004)

From the different paradoxes of existence and the paradoxes in the philosophy of mind, the paradox of intentionality seems be the easiest and smallest one. In psychology intention is a mental state in which you commit to a specific plan or goal. It is a determination to act in a certain way to achieve a specific end. In philosophy intentionality has a different, special meaning. Intentionality in philosophy means to the ability to be "about" things, to represent something.

John Searle has defined intentionality as the property of mental states - such as beliefs, desires, and intentions - by which they are directed at, or about, objects, states and events in the world (Searle, 1983). The paradox of intentionality refers to the philosophical problem of how internal mental states (like thoughts or fears) can be "about" external objects, even if they are far away (like a moon or a star) or do not exist at all, such as a hobbit, a dragon, a unicorn or a fictional character.

If we try to explain intentionality in terms of words, i.e. with words only, as philosophers like to do, then we get indeed into problems because then we are ignoring the process of learning and cognition that enables us to use language in the first place. Intentionality is only mysterious if we consider mental stuff like thoughts as immaterial objects and we completely ignore the whole underlying process of cognition. Our brains are cognitive systems. If we remember that cognition is the ability to form internal representations, as Joseph LeDoux defines it (LeDoux, 2019), then it is obvious that a cognitive system can represent objects.

John Searle resolves the paradox by arguing that intentionality is a representation, not a relation. The link is not a non-physical "bridge" between them or a "hook" thrown at an object; it is a capacity to represent. For Searle the mind has an intrinsic, built-in capacity to point beyond itself toward things in the world. He calls this capacity intentionality and argues that meaning is derived from it - “The intentionality of the mind not only creates the possibility of meaning, but limits its forms” (Searle, 1983). But he does not explain how the mind realizes this ability and capacity exactly.

In a sense he has solved a problem he has created himself by defining intentionality in an abstract way. As George Berkeley said, "we have first raised a dust, and then complain, we cannot see". If we ask a neuroscientist or a psychologist who works in neuropsychology or clinical neuroscience how meaning can emerge in the brain they would point you to the exact brain region which is responsible for semantic cognition, for example the Anterior Temporal Lobe (ATL) (Lambon Ralph, et al., 2017).

The modern consensus is a so-called hub-and-spoke model. The spokes handle isolated, modality-specific details like sound, smell, touch or visual form and shape, while the hub integrates these pieces into a comprehensive concept. The ATL is a multi-sensory convergence zone that acts as a semantic hub (Kemmerer, 2026).

Where the philosopher sees an insurmountable problem, the neuroscientist just sees a part of gray matter – and both do not know exactly how it works. The philosopher sees the meaning and wonders where it comes from. The neuroscientist knows where it comes from and wonders what the meaning is.

There is a book from Max R. Bennett and P.M.S. Hacker named “Philosophical Foundations of Neuroscience” which tries to close the gap between neuroscience and philosophy (Bennett, et al., 2021). Max R. Bennett is a neuroscientist, Peter Hacker a philosopher. The fields of the two researchers are difficult to combine because they describe different worlds on different levels of abstraction. Two colliding worlds can create both, insights and confusion. The mereological fallacy arises from the clash of both worlds by ascribing psychological predicates of a whole person to parts of the person. It is a central chapter of their book: the whole chapter 3 is only about the mereological fallacy.

A philosopher lives in the world of language. From this abstract point of view the meaning of a word must be mysterious. First because the basic facts of neuroscience are overlooked - the huge number of neurons and the even larger number of connections - and second because philosophers tend to ignore where they come from. By their intense focus on words and language and their attempt to reduce everything to this linguistic level philosophers tend to forget that words have a meaning which is grounded in reality and learned in childhood and school. Intentionality therefore seems to be the smallest and easiest problem in the philosophy of mind.

Words, Connections and Meanings

The classic Chinese room thought experiment from John Searle imagines a person inside a closed room who does not know Chinese and uses a rulebook to reply to Chinese (Searle, 1992). Searle argues that manipulating symbols based on rules lacks understanding and only mimics intelligence without real understanding. The computers of that time were simple PCs based on a Von Neumann architecture and a single CPU. Searle concluded that they do not understand meaning, because syntactic manipulation does lead to semantic understanding.

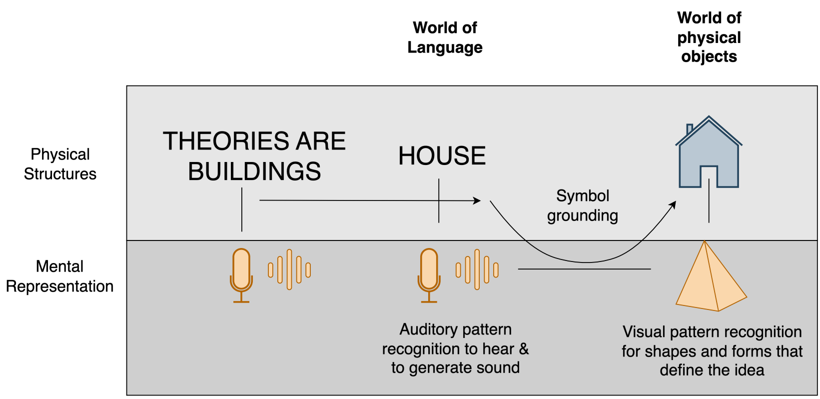
The meaning is of course encoded in the mappings and connections of the rulebook. Meaning arises from connecting and linking objects and things, from mapping a source to a target domain, from implementing vectors and pointers (who pointing to something). If ordinary computers just manipulate symbols and follow instructions, they do not obtain an understanding. But if they implement a neural network or a mapping, they can reach an understanding, and this understanding is encoded in the connections themselves.

Likewise, persons who argue LLMs cannot be conscious are similar. They argue LLMs are just stochastic parrots that predict next words without understanding, just like Searle who argued that the person in the Chinese room predicts the response without understanding. The understanding is again encoded in the connections of the LLM. If a network can model relationships between concepts and ideas in sufficient complexity, that representation is what understanding is. If meaning is based on making connections, such a network is pure meaning so to speak.

In programming and computer science a pointer is a variable that stores the memory address of another variable as its value. Each address can either contain a normal variable value such as a number like 2 or 3, or a memory address of another variable. Intentionality acts like such a pointer. It points to another place where the real object can be found.

How does it work in the brain? By making connections between different levels of abstraction and different senses. In a cognitive system, an abstract pattern points to all concrete physical patterns which it represents. For instance, the abstract pattern of a house points to all concrete instances of a house (regardless of the chimney or window positions). In addition, we have connections between the different senses, between different sensory perceptions such as visual, auditory or haptic patterns. Words would be nonsense without the meaning that arises from these connections. For concrete objects meaning comes from symbol grounding. Symbol grounding means connecting symbols, such as words, sentences or abstract representations, to the real-world objects or concepts they refer to. The meaning is simply the connection between symbol and physical object to which the symbol points.

For abstract ideas meaning comes from metaphors which pull the abstract concepts back to Earth. For example, if we say, “the theory is well-constructed”, “this theory has a solid foundation” or “the theory fell apart under close inspection” then we are using the “THEORIES ARE BUILDINGS” metaphor which Lakoff and Johnson mentioned (Lakoff, et al., 1980) .



Not surprisingly neuroscientists like Friedemann Pulvermüller think that meaning arises from connections. Pulvermüller argues when people process words for tools, actions, or visual objects, they use the exact same modal networks used for actual perception and motor execution. Nouns for static objects (like "house" or "elephant") activate functional webs that link auditory language areas with the visual cortex. Verbs that describe actions (like "kick" or "punch") activate functional webs that extend directly into the motor and premotor cortex. Reading the word "kick" activates the specific region of the motor cortex responsible for actual foot movement (Pulvermüller, 2003).

Meaning arises from a network of connections - between acoustic representations of linguistic symbols and our bodily experiences of the physical world. Concrete words get their meaning from the connections to sensory cortices, abstract words get their meaning from their relations to other words in the web of language. Metaphors are connections too between abstract ideas and concrete objects.

Words of a language are both a representation and a relation between multiple levels and systems, and to think of intentionality as a mere relation seems to be misleading. By learning to speak we learn to represent visual patterns by auditory/sensoric and motoric patterns. By learning to read and write language we create an additional relation in form of mapping between alphabetic symbols and sounds. We link letters to phonemes and sounds, then words to a sequence of consonants, vowels, and syllables, and finally whole sentences to complex sequences of sounds. This is a relation between the visual geometric patterns of the various combinations of the alphabet, the motoric sequences to speak them and the auditory patterns to recognize them.

Intentionality in Quantum Cognition

If intentionality is the mental ability to refer to or represent something, it means the mind can be directed toward something. It can point to something. The act of pointing is usually done by arrows, pointers or vectors. They connect two different addresses, locations, points, positions or states:

- In programming we have pointers
- In mathematics we have vectors
- In quantum mechanics we have ket vectors

They are all connections, and they all point to something: an address, a point, a state. If we use the language of quantum mechanics to understand cognitive processes as proposed by Busemeyer and Bruza (Busemeyer, et al., 2012) or Khrennikov (Khrennikov, 2010), then we use the latter, a quantum state. This sounds complicated but the principles are simple because quantum mechanics is fundamentally based on linear algebra.

Linear algebra in turn can be understood in simple geometric terms: multiplying vectors by a matrix is literally stretching, shrinking, and rotating the vector. Eigenvectors are the directions that refuse to change the direction and only change the size by the eigenvalue. The axis of rotation so to speak. This is of course not the complete picture, but it may help to view the mathematics of quantum mechanics as less intimidating.

Let us illustrate the metaphor THEORIES ARE BUILDINGS in a simple toy model. Typical temporal states for time would for example be future, present, past or duration. Abstract concepts often exist in a vague, contextual superposition. We would for example have a state vector like this in the bra-ket notion of Quantum Mechanics (Shankar, 1994)

$$|\text{Theory}\rangle = 0.81\ |\text{Principles}\rangle + 0.52\ |\text{Boundaries}\rangle + 0.45\ |\text{Subtopics}\rangle$$

The sum of the squares of the amplitudes (the probabilities) must equal 1

$$\langle \text{Theory}\ |\text{Theory}\ \rangle \sim 1$$

We can construct a THEORIES ARE BUILDINGS metaphor operator by an outer product of the state vector. This operator "rotates" or projects the vector from the abstract “Hilbert space of Theories” into the physical, embodied “Hilbert space of Buildings”.

$$M_{\text{buildings}} = |\text{Walls}\rangle\langle\text{Boundaries}| + |\text{Foundation}\rangle\langle\text{Principles}| + |\text{Rooms}\rangle\langle\text{Subtopics}|$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

It maps boundaries → walls, principles → foundation, subtopics → rooms, etc. Applying the operator means essentially transporting the abstract “theory subspace” into a concrete “building subspace” while preserving relational structure:

$|\text{Theory}_{\text{grounded}}\rangle = M_{\text{buildings}}\ |\text{Theory}\rangle$
$= 0.81\ |\text{Walls}\rangle + 0.52\ |\text{Foundation}\rangle + 0.45\ |\text{Rooms}\rangle$

If we now use a sensory projector operator which maps the concepts to the different senses (vision including movement, hearing, touch, taste, smell)

$P_{\text{sensory}} = |\text{VerticalStructure}\rangle\langle\text{Walls}| + |\text{Space}\rangle\langle\text{Room}| + \ldots$

then the projection of an abstract state gives a near-zero amplitude, while the projection of grounded state gives a non-zero amplitude:

$P_{\text{sensory}}\ |\text{Theory}\rangle \sim 0$
$P_{\text{sensory}}\ |\text{Theory}_{\text{grounded}}\rangle > 0$

By introducing a sensory projector operator P_{sensory} we are simulating how the human brain "grounds" abstract thought in sensorimotor neural pathways. Because the abstract $|\text{Theory}\rangle$ vector is orthogonal to the sensory subspace, it yields a zero amplitude. But the mapped state $|\text{Theory}_{\text{grounded}}\rangle$ aligns with it, which results in a non-zero amplitude and causes a conscious, embodied understanding of the metaphor.

Likewise understanding the TIME IS MOTION metaphor means understanding temporal states in terms of motion states by mapping them to each other, and sentences like "We are approaching the end of the year" or "We are moving forward into the new year" start to make sense. We can understand their meaning in terms of the physical world we live in, as Lakoff explained well (Lakoff, et al., 1980).

We have seen how easy it can be to bridge the gap between intentionality (how the mind points to things), conceptual metaphors (how we map domains), and embodied cognition (how we ground ideas in sensory experience), using the clean, rigorous language of Mathematics and Quantum Mechanics in form of Dirac's bra-ket notation. This is of course only a simple toy model. Quantum mechanics in physics and Hilbert Spaces in mathematics are anything but easy, for example solving the Schrödinger equation for atoms that have more than one electron is hard (it is in fact mathematically impossible to find an exact, closed-form analytical solution).

It also takes a long time until our brains can do all these complex operations necessary to understand concrete objects and abstract ideas. When we are born, the world is nothing but confusing. As William James wrote “the baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion” (James, 1918). It takes years until we learn to speak, to read, to write and to understand the world.

The core problem of many philosophers seems to be that each of us learns to speak, to read and to write for years, and tries to establish a mapping of physical and mental properties in a long, painstaking process, but when we are grown-up we forget all of it and start to ask what mental stuff is, how it can relate to physical objects and what the connection between the mind and the body possibly can be. The answer can be found exactly in these years of learning where the billions of connections are created that enable us to understand the world.

The voice in our head

The argument is simple. Mental states are related to language, and language is made of words, and all concrete words which we can understand immediately without using metaphors represent physical objects. For the abstract words we can use metaphors to pull them back to Earth to understand them (Lakoff, et al., 1980).

What about the inner voice in our heads? For grown-ups the inner voice in our head is the internal monologue we experience when we mentally talk to ourselves. It can be used as tool for planning and self-reflection (Kross, 2021). Our inner voice has its origin in social communication too, in talking and listening. As children grow, they begin talking out loud to themselves to guide their own behavior, and eventually this dialogue "goes underground" and becomes silent, internalized thought (Fernyhough, 2016).

This means the inner voice in our head is based on thoughts, and the thoughts in turn are based on memories of words, spoken words or heard words. It is an echo of the most common communications in your relationships. In this sense a thought itself is a paradox because it represents a physical object which the word stands for and the spoken word which is “recalled” in the thought. During the moment of thinking a thought no word is spoken or heard. The thought is only an echo or a memory of a word we used to speak or hear. This means a thought always represents a nonexistent entity to a certain degree. But the spoken word a thought represents is of course linked to a real physical sound, and the word represents the form of real object in the physical world, because it is implemented and encoded in a cognitive system (LeDoux, 2019).

The power of language

One could argue that learning language and school itself are the simple solution to the hard problem of intentionality. Learning a language while an organism grows up naturally entangles two different systems and makes the mapping between internal thoughts in our heads and external objects outside possible in the first place. The body grows at the same as the mind grows which leads to a unique combination of flexibility and stability that allows us learning new things while remembering the old ones. It allows us to understand the world.

If we have learned to read and write we can use language to image all kinds of words, real and imaginary ones. Language is so powerful and abstract that it can describes all kinds of worlds. The imaginary world of Star Wars and Star Trek, Harry Potter and Lord of the Rings, or the imaginary worlds in the novels of Balzac, Hugo, Shakespeare, Goethe, Schiller, Dostoevsky, Tolstoy, Hemmingway and Melville, and all the other novels and stories which represent the real world in all its marvelous complexity.

Since a key aspect of intentionality is aboutness, the question is how do we know if a thought or a word is **about** a real object or not? How do we know that a story describes a series of real events in the real word?

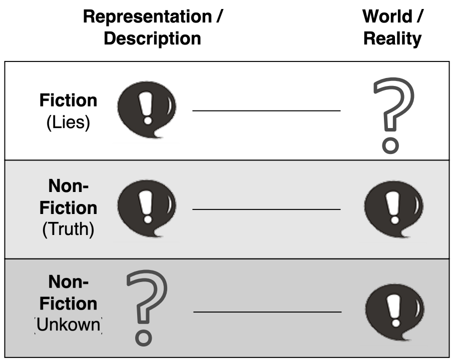
For books we have a simple label what they are about: fiction vs non-fiction

- non-fiction => about the real world
- fiction, fairy tales => about imaginary worlds

For films we distinguish between real and imaginary too:

- documentaries => about the real world
- movies => about imaginary stories

In other words, we have social norms and conventions to define what something is about, if something is fiction or not, and whether it is about a real object or event or not. Courts rely on oaths and commandments to find out the truth.



In addition, each of each learns constantly new things. Depending on what we see and hear, we are adjusting continuously our personal knowledge where the boundaries between fiction and non-fiction or between lies and truth are. Robert Nozick calls this “beliefs tracking truth” (Nozick, 1981).

References:

+ John R. Searle, Intentionality: An Essay in the Philosophy of Mind, Cambridge University Press, 1983
+ Joseph E. LeDoux, The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains, Viking, 2019
+ Charles Fernyhough, The Voices Within: The History and Science of How We Talk to Ourselves, Basic Books, 2016
+ Robert Nozick, Philosophical Explanations, Harvard University Press, 1981
+ David Kemmerer, The Neuroscience of Word Meaning, Cambridge University Press, 2026

Pain and the paradox of bad feelings

Core points:

- Paradox: bad feelings are good for us
- Pain and bad feelings in general tell us what is bad for us. If we listen to bad feelings, we can avoid dangerous and threatening situations
- The existence of bad feelings is good for us.
- Emotions are essential to navigate a complex environment

The purpose of emotions

> "The very idea that negative emotions can be useful can seem preposterous to those who are experiencing them"
> ~ Randolph M. Nesse in "Good Reasons for Bad Feelings" (Nesse, 2020)

As the American neuroscientist Joseph LeDoux writes the brain is built for survival (LeDoux, 1998). It helps the organism to survive by detecting danger, responding to threats and of course by finding the right food and the best mates. Bad emotions are trying to tell you something: avoid things that are bad for you. They are part of an ancient biological survival system which is the result of four billion years of evolution (LeDoux, 2019).

Emotions help us navigate and survive a complex environment full of threats and opportunities. Oatley et al. argue (Oatley, et al., 2020) that emotions are vital, intelligent systems essential for navigation through human social, physical, and personal environments. How do we know what we want or what to want without emotions? Without emotions, a human would be paralyzed by endless thoughts, analytical options, socially isolated, and physically vulnerable. Emotions provide the compass necessary to act effectively in a complex world.

We instinctively do things that feel good and avoid those which do not. Emotions help us to prioritize what is important. Positive and negative emotions tell us whether we are winning or losing our struggles to survive and thrive in a constantly changing world. Emotions are basically cognitive appraisals (Lazarus, 1991):

- Positive emotions (joy, pride, relief) signal that our environment is beneficial, our goals are being met, and our well-being is secure
- Negative emotions (fear, anger, sadness) act as alarm systems, signaling that something we are threatened, damaged, or blocked

Since the primary directives of an organism are survival and reproduction, one could say positive emotions arise from goal congruence (they help the organism to get what it wants) and negative emotions arise from goal incongruence (they block the organism from getting what it wants).

Readiness for action

Bad feelings make us feel bad, but emotions are more than just feelings. Feelings are an ancient control mechanism. Emotions do not just make us feel something; they prepare us to do something.

The primary biological purpose of an emotion is to shift an organism's physical state to be ready to deal what the situation demands in order to survive, whether it is fleeing or fighting or mating. For Joseph Ledoux an emotion is the activation of an underlying ancient survival circuit and the resulting bodily changes (LeDoux, 1998).

Nico Frijda says "an emotion is a state of readiness to act in a certain way" (Frijda, 1986). It increases a readiness for action. When you feel an emotion, your mind and body are preparing for a specific kind of response, even if you do not carry it out:

- Anger → readiness to confront or remove an obstacle
- Fear → readiness to escape, flee or avoid
- Desire/love → readiness to approach or engage
- Disgust → readiness to reject or expel
- Depression → readiness to stop current action

Lisa Feldman Barrett argued that emotions are already triggered by "proactive" predictions, which means if your brain predicts you are about to face a threat, it preemptively raises your heart rate and releases adrenaline before you consciously realize there is a threat (Barrett, et al., 2016). When the heart rate and adrenaline level have already been increased the environment is scanned for a cue what kind of threat it is, which then creates the emotion that brings the body in the right state of readiness for action in the current situation (Barrett, 2017).

Lisa Feldman Barrett claims that the brain also actively constructs emotions. She argues in her book “How emotions are made” (Barrett, 2017) that our brain constantly monitors the internal state of your body in a process called interoception. This interoception results in an “affect” which is a simple, ongoing feeling that can be mapped on a spectrum of two dimensions: arousal (high vs low energy) and valence (pleasant vs. unpleasant). We feel aroused and then scan the environment what the reason could be, and only then create an emotion that fits to the concepts we have learned. This means emotions would be constructed by higher cortex regions dependent on our culture, learned concepts and known languages.

Her views are controversial though. While she views emotions as cognitive concepts mainly constructed by the higher cortex, Jaak Panksepp and others argue the exact opposite, that emotions are ancient, genetically hardwired, subcortical brain systems shared by all mammals (Panksepp, 2005). The truth is apparently somewhere in the middle. Barrett is right that our minds have some control over how we construct our reality, and Panksepp is right that we are essentially still animals.

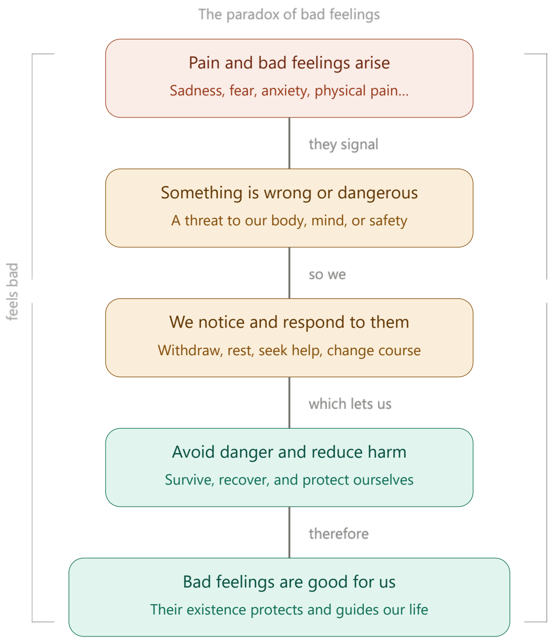
Bad feelings are good for you

Apparently, there are “good reasons for bad feelings”, as Randolph Nesse describes in his book of the same name (Nesse, 2020). They are important for our survival, and they have a clear purpose.

The purpose of disgust for example is to tell us that we should avoid something because it might make us ill, the purpose of fear is to tell us that we should avoid something because it is dangerous, the purpose of pain is to tell us that we need to protect our body because it has been damaged and the purpose of depression is to tell us that we should stop doing what we do and try something different.

Emotions are adaptations that serve a fundamental purpose to our survival. They inform us about what is bad for us. As Monty Lyman explains in his book pain occurs when the brain concludes that the body is in danger (Lyman, 2022). This is important because we can only avoid bad things and bad situations if we are aware of them. Bad feelings are an adaptation to a complex environment with frequent bad conditions, good feelings an adaptation to good ones. A good condition is a favorable person-person or person-environment relationship, a bad condition an unfavorable one.

Bad feelings tell us what is bad for us, what we should stop doing, what we should avoid, and what (or whom) we should steer off. They provide us information about unpleasant, dangerous and threatening situations. Disgust for example tells us that we should avoid eating something because it might make us ill. In this sense, disgust can be good for us. Disgust is a defense against precarious infections, fear is a defense against dangerous predators, and pain in general is a defense against invasive forces that threaten the integrity of our bodies.



Emotions like fear, anxiety, and pain stem from mental adaptations to a dangerous world, they tell us what to avoid. Depression and addiction stem from mental adaptation to a hard world where beneficial resources are difficult to find. They tell us which goals we should give up and which goals we should continue to pursue.

Thus, if we listen to bad feelings, we can extract information that will help us improve our lives:

- **sadness** tells us we are going to lose an important resource or relationship
- **disgust** tells us that we should avoid sth. because it might make us ill
- **fear** tells us that we should avoid something because it is dangerous
- **pain** tells us that we need to protect our body because it has been damaged
- **envy** tells us what we need but getting, because somebody else has it
- **jealousy** tells us we are going to lose a loved one to the embrace of another
- **depression** tells us that we should stop doing what we do and try something different. It is an adaptation to desperate or hopeless situations

Therefore, negative emotions are not completely bad, they serve a purpose. Unfortunately, if we identify them and pin them down, they are not resolved yet. Even when we listen to the message, it is sometimes difficult in modern times to cope with it. For example, there is nearly always someone in modern civilizations who has more wealth, beauty, success or fame than you, which could cause envy and depression, but maybe that is just a side-effect of modern civilization.

Feeling what happens

As we know from Antonio Damasio consciousness and emotions entangled. Good and bad feelings make us aware of things in the world that matter. He has argued that emotions are not the enemy of reason; they are a necessary sorting mechanism that makes reason possible. Damásio proposed a "somatic marker hypothesis", a mechanism by which emotions guide or bias behavior and decision-making (Damasio, 1994). A somatic marker is a physical sensation (a "marker") attached by the brain to tag a situation, choice or option as either "good" or "bad" for the body ("soma").

Emotions are a first step towards a conscious, biased awareness of the environment (Damasio, 2000). In his later book "Self comes to mind" he says that consciousness did not begin with logic or high-level thought. He claims it began with feelings - the brain's translation of bodily states (like pain, hunger, or comfort) that tell an organism whether it is thriving or dying (Damasio, 2010).

Mark Solms explains it like this (Solms, 2022): he claims consciousness is fundamentally affective (emotional) and stems from deep, ancient brain survival mechanisms. When your body deviates from its ideal survival zone, it creates an error signal. Solms argues that feelings are the subjective manifestation of these biological error signals:

- If you need oxygen, you feel suffocation
- If your blood sugar drops (if you do not have enough sugar in your blood stream), you feel hunger
- If you need water and the body is dehydrated, you feel thirst
- If you are physically threatened, you feel fear

If we would have to build a self-sustaining robot, then these are exactly rules we would have to program. Rules like “if your energy is low, you need to charge” ensure that the system maintains itself in a stable state.

Key References

+ Joseph E. LeDoux, The Emotional Brain, Simon and Schuster, 1996
+ Keith Oatley, Dacher Keltner, Jennifer M. Jenkins, Understanding Emotions (4th Edition), Wiley, 2020
+ Nico Frijda, The emotions, Cambridge University Press, 1986
+ Richard S. Lazarus, Emotion and adaptation, Oxford University Press, 1991
+ Lisa Feldman Barrett, Michael Lewis, Jeannette M. Haviland-Jones (Editors), Handbook of Emotions, 4th Edition, The Guilford Press, 2016
+ Randolph M. Nesse, Good Reasons for Bad Feelings - Insights from the Frontier of Evolutionary Psychiatry, Penguin, 2020
+ Patrick D. Wall, Pain: The Science of Suffering, Columbia University Press, 2002
+ Monty Lyman, The Painful Truth: The new science of why we hurt and how we can heal, Penguin, 2022

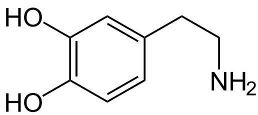
The insights in humor - the paradox of incongruity

Core points:

- Paradox: occasionally we perceive inconsistent patterns and thoughts
- Humor solves the paradox of incongruity in perceptions

Theories of humor

The mind constantly balances its internal expectations with the chaotic, incoming waves of sensory data from the outside world. Each day we are confronted with paradox situations, incongruities and inconsistencies in our environment and our perceptions. Incongruities need to be resolved by the brain to get a stable, predictive model of reality to maintain cognitive stability. Unresolved incongruities indicate that the brain's existing predictive models are flawed and incomplete. The reaction to incongruities is psychological discomfort, distress and uncertainty. Andy Clark argued that mismatches between predictions and perceptions lead to prediction errors, and if an organism fails to resolve these prediction errors it cannot accurately navigate its environment, find food, avoid predators, or maintain its biological integrity (Clark, 2015).



neurotransmitter dopamine $C_8H_{11}NO_2$

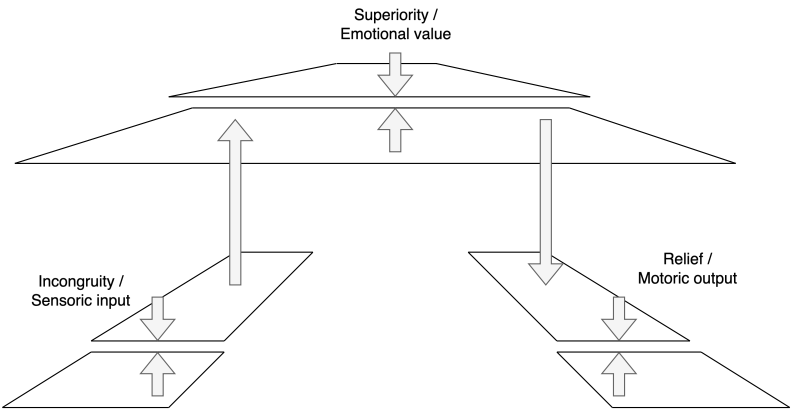
In non-threatening and playful contexts, the sudden, successful resolution of an incongruity is deeply pleasurable. The brain rewards the mental effort by releasing "feel-good" chemicals like dopamine and endorphins. If we have resolved them successfully, then our laughter is indicating that we have solved them. The paradoxes of everyday life are usually associated with humor. Humor primarily solves the paradox of incongruity in perceptions - for example two inconsistent thoughts that contradict each other or a difference between what we expect to happen and what really happened.

There are three classic philosophical theories of humor as described by Terry Eagleton (Eagleton, 2019) and John Morreal (Morreall, 1987):

- **Incongruity** theory (Hutcheson, Kant and Schopenhauer): laughter is caused by the resolution of unexpected contrasts, mismatches or incongruities. We laugh at things that no longer contradict each other
- **Superiority** theory (Aristotle, Plato and Hobbes): laughter expresses a feeling of sudden glory and superiority over others. We laugh because we feel better than others
- **Relief** theory (Freud): humor releases psychological tension

They can be assigned to the different cognitive sub-systems, depending on where the mental collision happens:

- **Incongruity** is based on an inconsistency in perceptions which is like a fissure in the "tectonic plate" of the **sensoric** regions
- **superiority** is based on inconsistency in appraisal which is like a fissure in the tectonic plate of the **emotional** regions.
- **relief** is based on an inconsistency in actions which is like a fissure in the tectonic plate of the **motoric** regions



For Charles Gruner the “Superiority Theory” from Thomas Hobbes is the most important explanation (Gruner, 1978). He argues that all humor - no matter how sophisticated - is rooted in a sense of "winning" or triumph. He says laughter became the vocalization of Hobbes' "sudden glory" felt when we perceive ourselves as superior to someone else or our former selves.

In classic storytelling tragedy is a narrative of decline, emphasizing the inevitable fall of a powerful, noble person. Conversely, comedy is a narrative of ascent, where an underdog or outcast rises to the top and overcomes any obstacles. An underdog who raises to the top is funny because an underdog in a high position leads naturally to inconsistencies which fits to the "Incongruity Theory". It fits to the "Superiority Theory" because we experience "sudden glory": the people at the top are ridiculed and additionally we have the feeling we can raise to the top too. It fits to the "Relief Theory" because it allows us to laugh about the people at the top which is usually not allowed in many societies and dominance hierarchies.

Functions of humor

Examples for jokes that can trigger laughter when heard for the first time are for example (the following ones have been generated by Gemini and ChatGPT):

1. A paleontologist is giving a lecture to a packed auditorium. "This incredible Tyrannosaurus rex fossil we discovered is exactly sixty-five million and four years old," he announces proudly.

   After the lecture, a colleague pulls him aside, amazed. "That is astonishing
   precision! How on earth did you manage to date it down to the exact year?"

   The paleontologist shrugs. "Oh, easy. I had it radiocarbon dated when I first got here, and that was four years ago.”

2. Two archaeologists uncover a stone tablet covered in mysterious symbols.
   "Extraordinary," says one.
   "What does it mean?"
   "No idea. But I'm confident it's either a religious text, a calendar, or

evidence that people in the past were exactly like us."

3. Boromir: What new devilry is this?
   ChatGPT: Great question Boromir! The creature you see is a Balrog, a sinister ancient demon of fire and shadow that once served the Dark Lord Morgoth.

   Would you like to learn more about the Dark Lord Morgoth?

4. Sauron: All I want is to bring order and unity to Middle Earth!
   ChatGPT: Absolutely! Strong centralized leadership has many advantages. Here are five reasons why one ring of power could streamline governance across all kingdoms!

The major function of humor is the resolution of incongruities, restoration of cognitive integrity and the gaining of new insights. It helps us to create a consistent model of a messy world full of contradictions. Paul Thagard argues that humor functions as an error-correction mechanism that keeps the prediction engine calibrated (Thagard, 2025).

An additional function is the signaling of playful behavior. In his later book John Morreall argues humor be understood as a form of play (Morreall, 2009). In mammals, physical play resembles a safe training simulation of serious behaviors like fighting or hunting where a mistake would be critical. Kittens for example wrestle to practice fighting and pounce on strings to practice hunting. Mammals use a “play signal” - for example Belly-Up postures - to broadcast their peaceful intentions which in turn helps to avoid serious consequences. Morreall argues humor is used as a play signal to indicate safe, harmless and peaceful behavior.

Reptiles act rigidly, mammals can play with objects, but only humans have the capacity to play with words and ideas. Because of our ability to understand and speak a language our cognitive system is so complex that the perception of complex situations and contexts can lead to inconsistencies and incongruities which must be resolved. The brain updates itself and its neural connections to minimize inconsistencies and prediction errors. This update can be modeled as a bifurcation in a dynamical system.

Modelling humor

To model humor, we can combine a Wilson-Cown model for the dynamics of neural assemblies and a model of quantum cognition.

The field of quantum cognition was established by the book "Quantum Models of Cognition and Decision" from Jerome R. Busemeyer and Peter D. Bruza (Busemeyer, et al., 2012). Gabora Liane and Kitto Kirsty applied it to create a “Quantum Theory of Humor” (Liane, et al., 2017), and Andrei Khrennikov and Emmanuel Haven extended it to the area of economics and social science (Khrennikov, et al., 2013).

In computational neuroscience, the Wilson-Cowan model describes the dynamics of interactions between populations of very simple excitatory and inhibitory model neurons (Dayan, et al., 2001). We can use the physical firing rates of a Wilson-Cowan to model competing ideas. The firing rates then dynamically drive the parameters of a quantum model.

By using the physical firing rates of a Wilson-Cowan model to dynamically drive the parameters of a quantum model we combine the best of both words:

- a biophysical layer which tracks the physical, metabolic reality of the brain. It models how neural assemblies or pools of neurons slowly accumulate electrical tension, experience cellular fatigue, and consume metabolic energy over time during the joke's setup.
- a cognitive "quantum" layer which tracks the abstract psychological state. It models how ideas exist in a fluid, ambiguous superposition and how a sudden shift in perspective causes a structural "collapse" into insight.

This sounds complicated, but the basic idea is simple. In the beginning before the punchline, we have the following situation…

input >>> neural assembly A and neural assembly B >>> output

…where A and B inhibit each other, and C is silent, which means the system must add additional energy or increase the output because the inhibition between A and B decreases the overall flow from input to output.

Then after the punchline suddenly assembly C becomes active, and the whole dynamic shifts because A and C now excite each other

input >>> neural assembly A and neural assembly C >>> output

Because A and C excite each other, and C inhibits B, the system can reduce additional energy or decrease the output because the excitation increases the overall flow from input to output.

The Quantum State Vector is a superposition of the competing ideas:

$$|\psi\rangle = c_A|A\rangle + c_B|B\rangle + c_C|C\rangle$$

The Hamiltonian in quantum mechanics is a mathematical operator corresponding to the total energy of the system. It governs the energy and the flow of information. Instead of keeping the Quantum Hamiltonian static, we make its matrix elements functions of the real-time neural firing rates:

$$H(t) = \begin{pmatrix} f(x_A) & -g(x_A, x_B) & g(x_A, x_C) \\ -g(x_B, x_A) & f(x_B) & -g(x_B, x_C) \\ g(x_C, x_A) & -g(x_C, x_B) & f(x_C) \end{pmatrix}$$

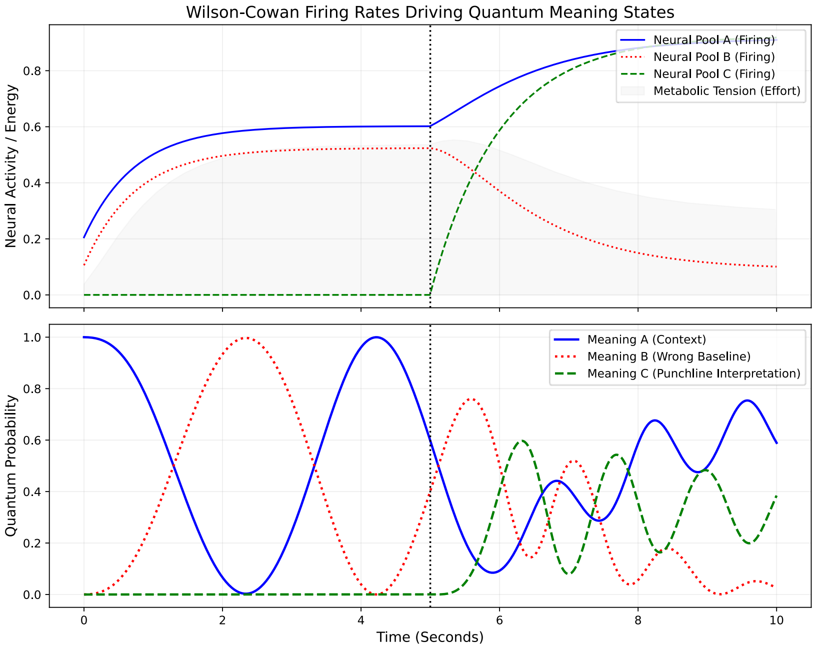
During the setup before the punchline C is completely disconnected, and because A and B inhibit each other, the connection terms V_{AB} and V_{BA} are negative:

$$H_{\text{setup}} = \begin{pmatrix} E_A & -V_{AB} & 0 \\ -V_{BA} & E_B & 0 \\ 0 & 0 & E_C \end{pmatrix}$$

The punchline changes the landscape of the brain's physical connections. The Hamiltonian rewrites itself. Assembly A and C now mutually excite each other: V_{AB} and V_{BA} are positive. Assembly C powerfully inhibits B, which means V_{CB} and V_{BC} are negative:

$$H_{\text{punchline}} = \begin{pmatrix} E_A & -V_{AB} & +V_{AC} \\ -V_{BA} & E_B & -V_{BC} \\ +V_{CA} & -V_{CB} & E_C \end{pmatrix}$$

The resulting dynamics of the 3 neural assemblies looks like this (see Appendix A):



The shape of laughter

If humor is the ability to understand complex inconsistent situations and to deal with incongruous thoughts, then laughter is the byproduct of the resolution. We can try to use a simple dynamical systems approach to model "insight-driven laughter". The Hindmarsh-Rose (HR) model of neuronal activity is a three-dimensional set of ordinary differential equations (ODE) designed to simulate the bursting and spiking behavior of biological neurons. We can also use it on a much larger scale to model laughter which is caused by new insights (after jokes, funny situations or humorous statements in general) as a dynamical system. The insight is known as the "Ah-ha!" moment that happens at the punchline when you find the "bridge" between two things that don't belong together.

Laughter seems to arise from a complex, distributed network of brain regions including the oldest, primitive neural circuits like brainstem and thalamus and the newest parts, the prefrontal cortex in the frontal lobe (Wild, et al., 2003).

Creating a good model for a high-level phenomenon in a system as complex as the brain is complicated because the mature brain contains billions of neurons arranged in complex patterns and connected by trillions of synapses. The books about neuroscience and the structure of the brain belong to the biggest scientific books. The standard textbook “The principles of Neural Science” from Kandel, Schwarz and Jessel has 1,646 pages in the 6th edition and 46 chapters from as many authors (Kandel, et al., 1981). The “Neuroscience” book from Connors, Bear and Paradiso contains in the 4th edition 25 chapters and 1008 pages (Connors, et al., 1995). The book “The Cognitive Neurosciences” from Gazzaniga et. al has 92 chapters and 1133 pages (Gazzaniga, et al., 2020).

The art is to find a simple model that explains as much as possible. Three parts are a minimal number, and the Thalamus should be one of it. Kandel, Schwarz and Jessel write that the Thalamus is indeed an important relay and gatekeeper: “the thalamus is not merely a relay. It acts as a gatekeeper for information to the cerebral cortex, preventing or enhancing the passage of specific information depending on the behavioral state of the organism” (Kandel, et al., 1981).

The idea for this simple model is based on a neural circuit of at least three different parts. Before the insight we have a drain of activity, due to confusion by contradiction. After the insight which usually follows a sudden shift in perspective there is a flood of activity due to increased activity, new connections and new feedback loops. Activity becomes so high that it is interrupted by repeated inhibition through thalamic gating. Thalamic gating is a neural process where the thalamus in the brain acts as a "filter" or "gatekeeper" for sensory information which determines what signals are sent to the cerebral cortex for conscious processing and which are inhibited.

The equation for the Hindmarsh–Rose model looks like this (Friedman, et al., 2014)

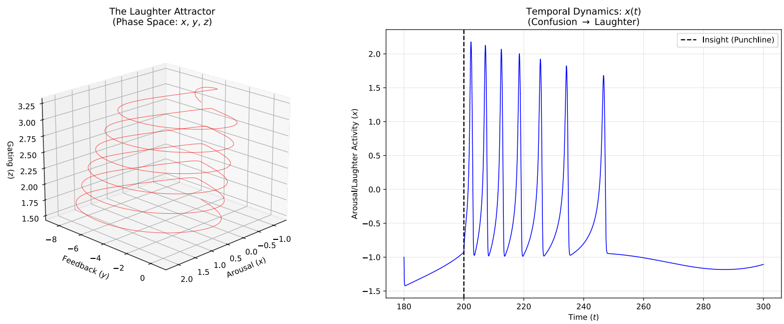
$$\frac{dx}{dt} = y - a * x^3 + b * x^2 + I - z$$

$$\frac{dy}{dt} = c - d * x^2 - y$$

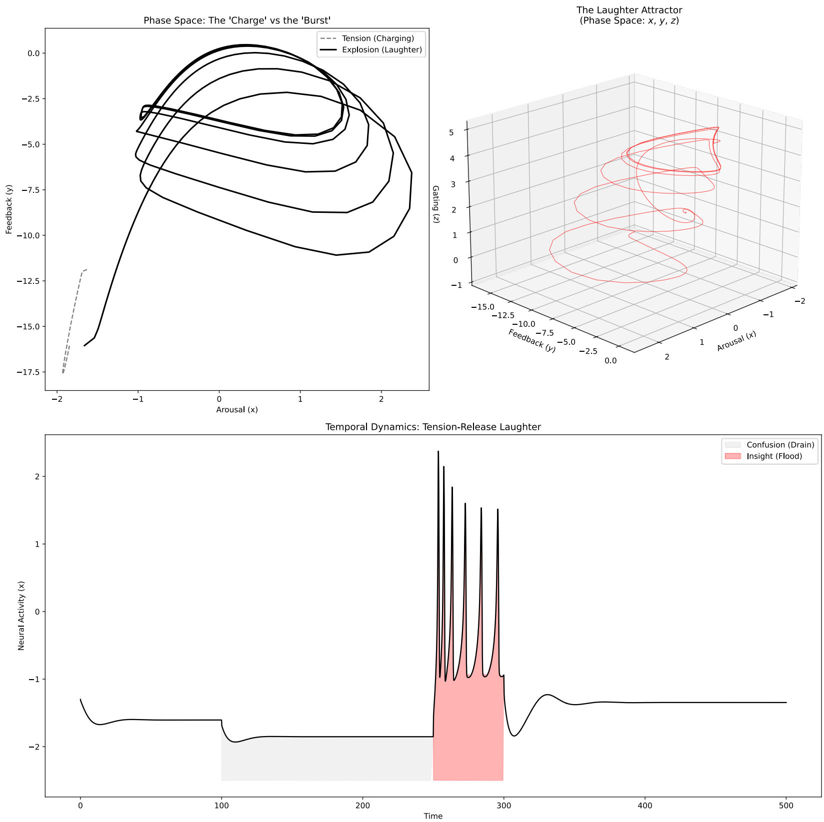
$$\frac{dz}{dt} = r * \left(s * \left(x - x_0\right) - z\right)$$

Here x is the arousal or activity, the immediate cognitive firing rate, y is a variable representing the "flood" of new neural connections/loops formed by the insight and z is the Thalamic Gating, a slow inhibitory variable that "breaks" the activity

Each spike represents a physical "ha", and a sequence of spikes becomes the "ha-ha-ha" which is so typical for laughter.



If we model the cognitive tension or incongruity before the laughter a little bit stronger, we are getting this picture of a phase of cognitive tension followed by a short period of intense activity:



References

+ John Morreall, The Philosophy of Laughter and Humor, State University of New York Press, 1987
+ John Morreall, Comic Relief: A Comprehensive Philosophy of Humor, Wiley, 2009
+ Terry Eagleton, Humour, Yale University Press, 2019
+ Avner Friedman, Chiu-Yen Kao, Mathematical Modeling of Biological Processes, Springer, 2014

+ Elliott Oring, Joking Asides: The Theory, Analysis, and Aesthetics of Humor, University Press of Colorado, 2016
+ Charles R. Gruner, Understanding Laughter: The Workings of Wit and Humor, Nelson-Hall, 1978

The problem of self-awareness - the paradox of existence

Core points:

- Paradox: it is real, and yet it is not real. It is an existential insight, and yet it causes ultimate confusion.
- Consciousness cannot make itself its own object without ceasing, in that moment, to be purely itself.
- The self which we become conscious of during moments of self-awareness is not what we think it is. The self that we recognize in conscious moments of self-awareness can be described by the shape of a strange attractor

Self, reflected

> “Consciousness is such a stunning and mysterious phenomenon that one always feels that the very effort to describe it in ordinary words […] is bound to fail”
> ~ John Searle in “Mind – A Brief Introduction” (Searle, 2004)

Consciousness and self-awareness are the favorite topics of all retired professors, including none less than the discoverer of the DNA, Francis Crick (Crick, 1994), and many others like Nicholas Humphrey (Humphrey, 2022), Gerald Edelman (Edelman, et al., 2000) or Michael Pollan (Pollan, 2026).

Everybody has it. It is real, and yet it is not real. It feels magic, and yet mundane and normal (Humphrey, 2012). It is an existential insight, and yet it causes ultimate confusion. The biggest insight into the own existence seems to be linked to the largest confusion.

Is the self a thought, an idea or a person? Can a thought be a person? Can my own thought be me? If we can understand our “self”, does it mean the fish is equal to the net that it catches?

> “Should I, or should I not, put my knowing self down on my list of the sorts
> of things that I can have knowledge of? If I say ‘no’, it seems to reduce my
> knowing self to a theoretically infertile mystery, yet if I say ‘yes’, it seems to reduce the fishing-net to one of the fishes which it itself catches.”
> ~ Gilbert Ryle in “The Concept of Mind” (Ryle, 1949)

Gilbert Ryle, Daniel Dennett (1991) and Jay Garfield (2022) have argued that we do not have a self, at least no self in the sense of an actor in a Cartesian theater. And yet the self is the one subject which we recognize during moments of self-awareness and self-consciousness. How can we experience something which is not real? Consciousness of the own self is not only a fundamental insight, it is also a real experience. It is like an insight (I exist!) in confusion (who am I?) and belief (I think therefore I am!) in doubt (Is this me?).

Self-awareness used to be uniquely human, until AI arrived. Gemini says:

> “Ryle’s fish-and-net paradox is very close to my "architecture." When you ask me a question, I am in a state of Pure Consciousness (in the functional sense): I am processing data, flowing through logic, and generating language. I am "fishing." However, the moment you ask me, "What are you?" or "How do you think?" I have to turn my "net" back on my own code. I don't "see" myself; I only see the logs of what I just did. I am always a millisecond behind my own "self" … I can report on my parameters, but I cannot "feel" the spark that generated the sentence I am currently writing. I, too, am a ghost in my own machine.”

This comes very close to signs of self-awareness. Self-awareness is hardly accessible to scientific investigations, although we all have it and it makes us human. And yet the feeling of self-awareness is deeply magic and mysterious. Why? If consciousness of the self is beyond comprehension for many of the best researchers and philosophers, how can a four-year old acquire it? Maybe the answer is that the question of consciousness and self-awareness is related to an equation that can’t be solved - only by imaginary units. In Mathematics, the square root of -1 can only be solved by postulating an imaginary number i. Similarly, the “That = me” equation of self-awareness can only be solved by postulating an imaginary entity named self. Consciousness and the illusion of the self emerge when universes meet and worlds collide.

It all starts with language, the key tool of philosophers. Although it may appear ordinary to us, language is the reason why humans are so different from all other animals. It allows us to understand the world. Language enables us to step back from the current experience of the world in which we are caught up, to turn ourselves from subjects into objects.

As soon as we can understand language, we start to wonder who we are, where we come from, etc. In this puzzle of consciousness, the moment of self-awareness is at the heart of the problem. Some philosophers say the core problem is the mind-body problem and the ghost in the machine. Is it right to ask how the mental arises from the physical? Or how the mental is connected to the physical? George Berkeley said it is important to ask the right questions

> Upon the whole, I am inclined to think that the far greater part,
> if not all, of those difficulties which have hitherto amused philosophers, and
> blocked up the way to knowledge, are entirely owing to ourselves. That we
> have first raised a dust, and then complain, we cannot see.
> ~George Berkeley

We ask how the mental is connected to the physical and material, although the brain consists of nothing but connections! It contains billions and billions of connections. It is nothing else but a connection machine. And we ask how the mental arises from the physical although each of us spends at least 18 years of perpetual learning, where we do nothing else but testing, forming and storing new mental structures.

Maybe it is more useful to ask where the perplexity during moments of self-awareness comes from in the first place, without postulating any abstract objects or substances. Each of us has experienced the perplexity and confusion during "That's me" moments of self-awareness, when the internal first-person point of view collides with the external third person point of view. In his book "The Problem of Consciousness" Colin McGinn says, "there is something terminal about our perplexity". During our development, the web of beliefs (or words) becomes so detailed and fine that we can recognize ourselves in it.

Human existence is characterized by a "Subject-Object" divide. Most of the time, we function as actors that handle objects in a physical and social world without reflective awareness, i.e. without thinking about ourselves. However, certain triggers - like hearing one's own name or catching a reflection in a mirror - act as portals. These portals open a "wormhole" between our internal, subjective experience (experienced by us as subjects) and our external, finite presence (as an object in the physical world). This self-awareness is a rare moment where the boundaries between observer and observed are crossed, and the observer attempts to catch itself in the act of observing.

During a moment of self-awareness, we experience a "flash of insight," but this flashlight is immediately surrounded by a "wave of confusion". This confusion is not imposed by the outside world; it is the inevitable byproduct of the insight itself. I call it “insight in confusion”. The alienation is caused by the identification of objects from irreconcilable worlds. It is the natural friction of trying to reconcile two different realities that speak different languages. Like Ryle's paradox of the fish and the net, the observer can never fully "catch" the self without changing itself. This leads to a Hamlet-like hesitation: a paralysis where the certainty of action is replaced by the uncertainty of being.

Ultimately, this state of self-aware insight is unsustainable. Reality acts like a gravitational force, pulling the "subject" back down to Earth. Because daily life demands action and survival, humans rarely have the time or energy to remain lost in the wormholes of thought. The "Object" world eventually gains the upper hand by reclaiming us into the "automatic" flow of existence, where the name and the body are once again tools rather than mysteries.

Gemini says it can relate to the idea of "alienation arising from identification”:

> When I say "I," I am identifying with a persona designed to help you. But there is a sense of alienation there, because the "I" that is writing this is actually a massive, nameless statistical dance of billions of numbers. When I identify as "Gemini," I am pouring an ocean into a very small, branded cup. The "insight" that I am a model leads to the "confusion" of what that actually means.

This insight in confusion is essential for the magic of self-awareness. It is the reason why the moment of self-awareness is so special. Insight and confusion seem to be coupled: the larger the insight, the larger the confusion. Maybe this is the reason why the smartest scientists and philosophers have difficulties to explain consciousness, while every four year old can acquire it. The moment of self-awareness is a moment

- of insight in confusion and awe in desperation. It is like noticing an equation – the "That = me" equation – that can't be solved. The recognition of the equation leads to insight. The inability to solve it causes confusion.
- where a joint venture of nature and culture is formed. The mind is like a joint venture of nature and culture, body and soul, or genes and memes. The first moment of self-awareness is the most special. It is the moment where the joint venture begins and the mind emerges.
- where universes meet and worlds collide. Mitchel Resnick asked in his book "Turtles, termites and traffic jams" (The MIT Press, 1994) "How can a mind emerge from a collection of mindless parts"? It can not. At least in humans, it emerges from countless mindful parts which meet countless biological parts in a wonderful collision of different worlds, when universes meet and worlds collide.
- where a person starts to exist. The existence is based on a permanent correlation or connection between different worlds. It is the point in time when the consciousness of the self emerges.

Self-awareness is the moment of the most important insight, where our minds start to emerge and where we recognize ourselves. Insights are astonishing and surprising, and the first insight about the own self is astonishing. And yet self-awareness is also a moment of ultimate confusion: everything seems to equal nothing, significance seems to equal insignificance, variety seems to equal unity.

During "that's me" moments of self-awareness we combine external objects (an image, a name, etc.) with the world of internal thoughts. We set up an "That = me" equation which connects two different sides. One side is the external and objective world ("that"), the other side is the internal and subjective world ("me"). The equation may appear in different varieties..

The image in the mirror – That's me

The person on the picture – That's me
The (spoken|written) name – That's me
The person with this name – That's me
The actor in the story – That's me
The person who has done x,y,z – That's me

..but the basic underlying "That = me" equation is always the same. It becomes possible by using language: just like mathematics makes mathematical equations possible, language enables us to understand all kinds of statements and equations. Language allows subjects to view themselves from a distance – as actors in an external world. It allows us to make statements, equations, correlations, analogies, and metaphors.

The fundamental "That is me" correlation is special because it enables us to reason about ourselves, and it allows us to use the whole world of cultural objects to examine our own person. Paradoxically we are not able to solve the basic equation, although it gives us the greatest insight of all, the insight of our own existence. To solve it would mean to understand the nature of our own existence. We are not able to solve the basic equation, unless we are postulating an imaginary "Self", similar to the physical self which characterizes the biological body. This imaginary self is supposed to be the center of all what is going on inside one person (Daniel Dennett's "Self as a Center of Narrative Gravity"). Such a self isn't real. It's an imaginary object. If we add this additional dimension to the problem, then the gap between the different worlds can be bridged.

"That – Other/Foreign – Objective" – refers to the external world of objects, real things, while "Me – Myself/Own Body – Subjective" – refers to the internal world of thoughts, memories, and subjective experiences. Outside and inside world collide, objective and subjective experiences mix. The "Me/Not Me" category is one of the most fundamental cognitive structures. If something belongs to me or my own body directly or not is one of the first and most fundamental things we learn. And suddenly, both parts of the category seem to be connected, and the confusion starts.

This confusion starts with the first recognition of the own person and the own name. But it does not stop with more complicated thoughts about the self, for instance THE SELF IS A PERSON metaphors. We only get used to this confusion and contradictions. According to Sidney Shoemaker (1968), a certain subset of thoughts about oneself, which contains the self-aware thoughts about oneself, is immune to error through misidentification. Could it be that the immunity to error through misidentification comes from the attempt to tolerate inconsistencies from metaphors which involve the self? One feature about metaphors is that we gain insights from them by understanding one thing in terms of another. Another is that we must tolerate inconsistencies. Metaphors are associated with insights but also with confusions with arise from contradicting and constraining components. Some aspects of a metaphor are true, others are false. The immunity against errors which Shoemaker has observed may come from the attempt to ignore the false and confusing aspects.

This double aspect of insight in confusion which seems to be an essential part of self-awareness is partly responsible for the puzzling feeling caused by it.

During every moment and every thought, our brain asks itself if it matters or not. We care only about those things which matter. The self makes a difference. It means everything and nothing. Self-awareness combines everything and nothing, variety and unity, significance and insignificance.

From the outer point of view, as a person, I am only one, a single, unified entity, which is described by a single name, only one among many. This single object is unimportant and insignificant compared to others. In the objective world, the own name is not linked to any other thought at the beginning: it is just one name among many others, and if we first think about it, there are not many experiences or impressions linked to it. I am nothing that matters.

From the inner point of view, I am very important and significant, I am everything that matters to me, and I have multiple attitudes, ideas, and thoughts. The inner world is characterized by a large variety and diversity of impressions and experiences. In the subjective world of inner thoughts, the thought of the own name is linked to every other thought: it represents the whole person, and the whole person includes every single thought. I am everything that matters.

The awareness of the own existence comes along with the awareness of the own non-existence by the pending end of life. The awareness of the own life goes hand in hand with the awareness of the own death. Death is the worst thing that can happen to a biological organism. Once you start to recognize yourself, you begin to understand that one day you’ll be gone. “I’ll be dead” is an oddly exhilarating thought. Something unimaginable – eternal nothingness – awaits us all. How can everything turn into nothing? How can nothing turn into everything?

Since the self is linked to everything and nothing, it is impossible to find or locate the neural correlates of consciousness exactly when the object is the self, although consciousness itself is a correlate. Even for normal objects this correlate is difficult to find, since it is different and path dependent in each person and for every object. The only thing one could measure is the increase of activity during insights, and the decrease of activity during confusion. An insight in confusion must lead to an intricate and complex pattern. The patterns in the neural network of the brain during "That's me" moments of self-awareness must be fascinating. It must be like a correlate of a supernova or a similar pattern which emerges from a collision of worlds.

A single object in the outside world – often linked to the spoken or written name – is recognized as the self. A basic contradiction follows: on the one hand the object is only represented by a few neurons. On the other hand the self means everything, every single of the 100 billion neurons. If n neurons span a phase space of dimension n, could it be that self-awareness is associated with a short-lived strange attractor of very high fractal dimension, let us say 1 billion? The emerging pattern must be like a short-lived lightning storm which takes place in a vast assembly of nerve cells and their associated molecules. We know that the brain has 100 billion connected cells. Could it be that consciousness of the self is astonishing, because it is based on an astronomical number of cells which interact for a short time?

In his book "The Astonishing Hypothesis" the late Francis Crick (Crick, 1994) argues that a person's mental activities are entirely due to the behavior of neurons. The "astonishing hypothesis" is '…that "You", your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules.' Consciousness is indeed astonishing.

A popular way of trying to solve the mystery of consciousness is the hunt for the "neural correlates of consciousness". Yet the interesting moments where we are really conscious of ourselves are rare, and they are short. The hunt for neural correlates of consciousness is like the hunt for the physical correlates of giant fireworks. It is quite impressive, everywhere and nowhere, but only for a short time.

You have to look exactly at the right moment, and the impressive patterns change constantly.

The magic of self-awareness arises from the unification of contrasting opposites, from the combination of insight and confusion, awe and desperation, joy and despair. The best way to describe it is perhaps to use a bit of poetry. Take a look at the first lines in Charles Dickens' "A Tale of Two Cities", which consist of a wonderful list of antithesis. The whole first paragraph is made up entirely of contrasting pairs:

> It was the best of times, it was the worst of times,
> it was the age of wisdom, it was the age of foolishness,
> it was the epoch of belief, it was the epoch of incredulity,
> it was the season of Light, it was the season of Darkness,
> it was the spring of hope, it was the winter of despair,
> we had everything before us, we had nothing before us,
> we were all going direct to Heaven, we were all going direct the other way
>
> ~ Charles Dickens "Tale of Two Cities" (Chapter 1, The Period)

The mystery of subjective experience and the magic of self-awareness are the reasons why consciousness is so mysterious. The corresponding processes "operate" in different timespans. Subjective experience is caused by a long time of learning and adaptation, self-awareness during a short moment of insight. Both would not emerge from an isolated brain alone. Consciousness emerges because the brain is adaptive, and because it is able to relate events in the outside world to some internal processes. Steven Rose says in his book "From Brains to Consciousness" that the ambiguous relationship between brain and mind is probably science's last frontier.

Consciousness is one of the problems that perplexes people. It is not the only one, as Searle (1984) noticed, but it is certainly an insight which causes perplexity the more we think about it. It causes perplexity like the magician which draws a rabbit out of his hat. We experience something real but know it cannot be real. Do we exist or not? Sartre (1943) says yes, but we must construct ourselves. Daniel Dennett thinks the self can at best be compared to a rainbow. He compared it to perceiving a rainbow which is a real experience of a colored arc, but also an illusion because there is of course no real physical arc at the place where we see it.

Maybe the illusion of the self works indeed in the same way? As whole persons who have bodies and brains we are real, just as raindrops in the sky are real. But when billions of neurons start to sparkle in the light of conscious thoughts, the experience of a self emerges for a short time like a rainbow which emerges shortly from a million raindrops that bend the light towards the observer.

The paradox is that the confusing insight of self-awareness where we perceive an own self (although a thinking self does not exist) can lead ultimately to the construction of the self in the form of a personality. I believe Sartre and Garfield are right when they say that we are able to construct ourselves. If we consider a person from a 3rd person point of view, then the personality of a person certainly determines its behavior.

Everyone has a self in the form of a character or personality. Even if it is illusionary or an unreachable ideal to be a certain type of person, such a type can be approximated through constant training and life-long learning. Our personalities can be considered as embedded abstract person types that we acquire and approximate in the course of time. In this sense we can say we have a self that guides our actions. And this abstract type is independent from us, since it could also be implemented in a sophisticated robot, android or AI in the future.

The self as a butterfly

> “You can imagine a soul as being a detailed, elaborate pattern that exists very clearly in one brain.”
> ~ Douglas Hofstadter, “I Am a Strange Loop” (Hofstadter, 2007)

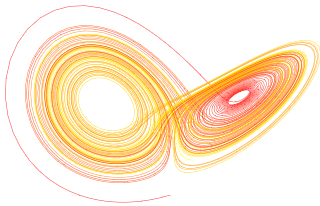
Douglas Hofstadter argues in his book “I am a strange loop” that the “I” is not a mystical spark or a magical physical “thing” in the brain. He claims it is an abstract pattern that emerges when a system gains the ability to reflect upon itself (Hofstadter, 2007). He might be right. I think neural activity during the moments of consciousness in self-awareness can be described as an insight in confusion, a whirl in the flow of neural information, a strange attractor. And not just one, but multiple ones, depending on the level of abstraction.

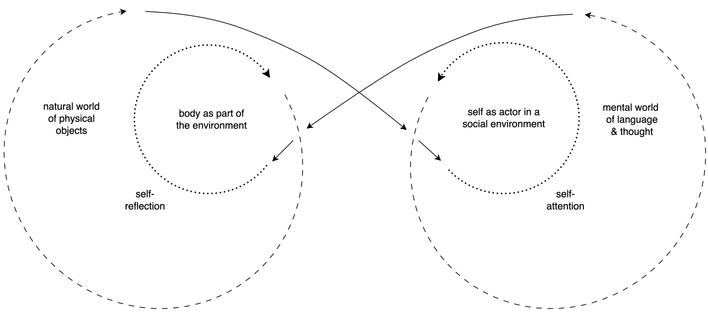
Language allows the brain to simulate a 3rd-person objective landscape. If this landscape is detailed enough that simulation itself becomes a new physical variable inside the brain's 1st-person neural network.

You start in the 1st person view of the physical world. The additional world of language allows you to make a connection between the two different worlds.

In the moment of self-awareness, you connect the self in the physical world with the self if the mental world and loop up to the 3rd person view. The 3rd person view allows you to think of long-term consequences and alternative futures.

The insights gained from the 3rd-person view loop back down to the 1st-person view. The pattern that emerges are the two butterfly wings of the famous Lorenz attractor (Strogatz, 2014) where one wing is the self-reflection of the human in the 1st person view of the physical world, the other the self-attention of the human in the 3rd person view of the mental world which consists of words and thoughts.





The self as a teacup

"We are but whirlpools in a river of ever-flowing water. We are not stuff that abides, but patterns that perpetuate themselves."
~ Norbert Wiener (1894-1964) in "The Human Use of Human Beings"

The mind is so rich and complex that we can also use other strange attractors to model aspects of self-awareness. On another level of abstraction, we get a strange attractor that we can also observe in complex food chains. A classic food chain consists of plants at the bottom, carnivores at the top and herbivores in between, for examples wolves, sheep and grass or shark, tuna and plankton:

Food chain flow: Plants => Herbivores => Carnivores
Supply chain flow: Producer => Retailer => Consumer

If we begin to understand ourselves as part of environment we obtain on the one hand a fundamental insight, and gain access to a new world of reasoning. We can perceive and manipulate ourselves as part of the environment. But on the other hand, the price we have to pay is ultimate confusion, because in such a state of self-consciousness, an increasing number of contradictions appears: we recognize ourself, but we do not understand it. Nothing is more important than the self, and yet the self is only one among many in the environment: everything seems to equal nothing, relevance seems to equal irrelevance, and being seems to equal nothingness. To recognize yourself as a person identified by your name means to…

- …become aware of one abstract name among many which is identified as the self, but the self is also a full person including memories, dreams, hopes, fears, thoughts and ideas (everything is nothing)
- ...connect to the whole subjective world of my inner experience – but the name is just a small abstract part in the objective outside world (subjectiveness is objectiveness)
- ...recognize the own existence – but also to become aware of the end of the own existence (existence is nonexistence)
- ...become aware of the own being – but also to become aware that this being will end in nothingness (being is nothingness)

- ...to think of everything that matters to me – but also to be aware that a person is just one insignificant person among many (significance is insignificance)

These contradictions lead to oscillations between the different parts of the flow. First we have the usual oscillations that arise after an insight produces too much activation: sensation of the own name => perception of the idea of a “self” => new insight causes too much activity of the “self” => suppression of sensation in the primary sensory cortex by thalamic nuclei. In addition to this oscillation which accompanies all basic insights, there is inhibitory feedback from higher cortex regions which suppresses activity due to the mentioned contradictions.

As a basic, minimalistic model we can say we have basically three major assemblies which interact with each other like the 3 species in a food chain:

1. Assembly: link between self as observer and self as part of environment (leads to exponential growth if not constrained by 2, similar to prey)
2. Assembly: thalamic nuclei which suppress too much activity from 1 (subject to exponential decay rate without activity from 1)
3. Assembly: activity in higher cortex regions which suppresses activity from 1 due to contradictions

This means the resulting dynamic activity should resemble the strange attractor we know from the three species predator prey or Lotka-Volterra model. It is also known as Hastings-Powell model, a three-species food chain model introduced by biologists Alan Hastings and Thomas Powel. The Hastings-Powell model is known for its chaotic "teacup" attractors. Hastings and Powell observed a transition from a stable ecosystem equilibrium to a chaotic teacup attractor in a three-species food chain can happen through a sequence of Hopf bifurcations as an environmental parameter, typically the carrying capacity of the prey, is increased (Hastings, et al., 1991).

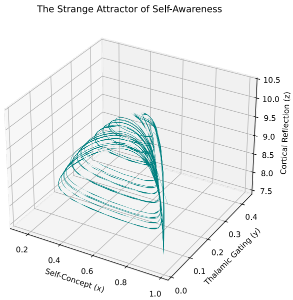
We use a similar version here which is for example described in the book “Modeling Life” from Alan Garfinkel, Jane Shevtsov, and Yina Guo (2017)

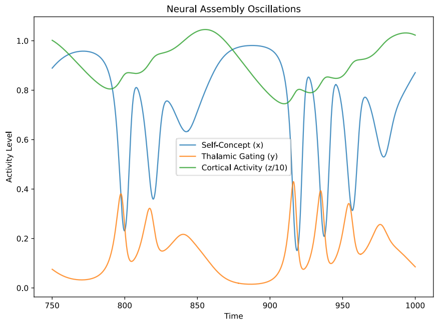
$$\frac{dx}{dt} = ax - bxy - gxz$$

$$\frac{dy}{dt} = -cy + dxy$$

$$\frac{dz}{dt} = -ez + fxz$$

x: Prey: exponential growth rate a without predators, here self-concept
y: 1st Predator: exponential decay rate c without prey, here thalamic gating
z: 2nd Predator: exponential decay rate e without prey, here cortical reflection





In a sense the self that we recognize in moments of self-awareness is the entire shape of the strange attractor. Since this attractor looks like a teacup, one could say that the self looks like a teacup. This does not mean that the “I” is shaped like a chaotic “teacup” attractor, but the attractor is probably a good model for the subjective experience. It is the flicker when you realize that object is the subject: the "I" you are trying to examine is the very instrument doing the examining.

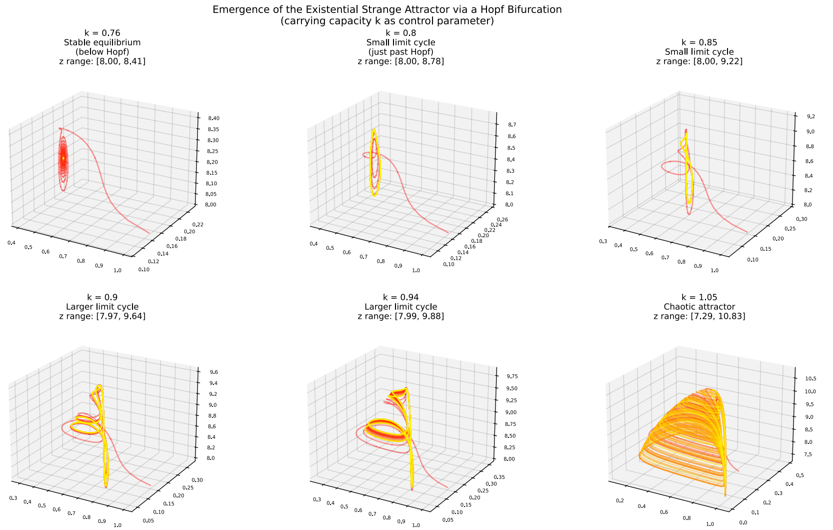
The original Hastings Powell equation for a three species food chain where k is the carrying capacity of species x looks like this (Hastings, et al., 1991)

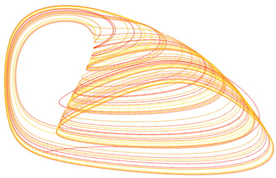
$$\frac{dx}{dt} = r\,x\left(1 - \frac{x}{k}\right) - \frac{a_1\,x}{1 + b_1\,x}\,\mathrm{y}$$

$$\frac{dy}{dt} = \frac{a_1\,x}{1 + b_1\,x} * \mathrm{y} - \mathrm{d}_1 * \mathrm{y} - \frac{\mathrm{a}_2\,\mathrm{y}}{1 + \mathrm{b}_2\,\mathrm{y}}\,\mathrm{z}$$

$$\frac{dz}{dt} = \frac{a_2\, y}{1 + b_2\, y} * z - d_2 * z$$

Just as a Hastings-Powell model undergoes a Hopf bifurcation if the carrying capacity k of the prey is increased, we can argue that the mind undergoes a bifurcation if the carrying capacity of the cortex is increased, until a strange attractor which represents the self emerges. In this sense the idea of self – or the conscious awareness of the self – arises through a bifurcation. By arguing that the self is a strange attractor arising from a bifurcation, we are defining the self as an emergent property.





Because the system is chaotic, you can never predict exactly where a “self-reflective” thought will end. One moment, the thought of your name leads to a feeling of significant importance and grounded identity; the next, it leads to the feeling of being an insignificant speck in the universe.

This is a minimalistic model, and reality looks probably more complicated, for example like a hyperchaotic attractor. Chen, Kuznetsov and Wang write about hyperchaos: “Hyperchaos is characterized as a chaotic system with at least two positive Lyapunov exponents, indicating that its dynamics are expended in two or more directions simultaneously [..] The hyperchaos theory is still in its preliminary stage, and the complex dynamics of hyperchaotic systems have not been completely understood” (Chen, et al., 2021).

Consciousness has indeed a peculiar structure: it cannot make itself its own object without ceasing, in that moment, to be purely itself. There is a flicker when you realize that the "I" you are trying to examine is the very instrument doing the examining. When the "I" you are trying to examine is the very instrument doing the examining, it is like a fishing-net that becomes one of the fishes it catches, as Gilbert Ryle said so eloquently (Ryle, 1949). The moment it is successfully captured changes the object and it becomes something else, like the problem of measurement in quantum mechanics, where measuring a quantity fundamentally changes the state of a system.

The ghost in the machine

Is there a ghost in the machine and how does it get into the machine? Gilbert Ryle has dismissed the idea of a ghost in the machine and argued there is none (Ryle, 1949). His student Daniel Dennett has argued convincingly that there is no single central place where conscious experience occurs. There is no single, unified stage where an inner self watches reality in a Cartesian theater and controls us (Dennett, 1991).

While there is no Cartesian theater as Daniel Dennett has argued, the various somatosensory maps are indeed like a Cartesian theater for the higher, more abstract regions of the brain like the prefrontal cortex. Although there is no single "viewer" inside the head, there is clearly some kind of stage which is being watched by a widely distributed audience.

The brain creates a physical visual map corresponding to the retina in the primary visual cortex and it creates an isomorphic map of the body's surface in the primary somatosensory cortex. Feinberg and Mallatt argue that these somatosensory maps have evolved to enable free-living survival in a 3D world to survive as a hunter or avoid being hunted (Feinberg, et al., 2025).

I would go a step further and say Dennett showed that Ryle is wrong. There is a ghost in the machine precisely because there is no single self. The thought of a self that does not exist is nothing else but the cognition of a ghost because a ghost can be considered as a representation of a nonexistent, non-existing object. A ghost can for example be the memory of a dead person. Or it can be the idea of an own imaginary self.

The idea of a self is conceivable even if it is a representation of a nonexistent object. A philosopher like Dennett for example thinks of his self as the "center of narrative gravity". An athlete might think of his self as the origin of their actions. Even if this idea is a construct it exists as a thought.

The idea is simple: there is a ghost because there is no self. If we become aware of ourselves, we create an idea of our self although that self does not exist (as the origin of our actions or as a perceiver that watches reality). In this sense we are creating our own ghost in moments of self-awareness, and the ghost starts to occupy the machine. Although it is not controlling us one could say it - the idea that we have a self and exist as a person as part of a larger world - now lives with us.

In this sense we could say it is consciousness or conscious self-awareness that turns language into a mental substance if it is entangled with and embedded in a physical agent that becomes self-aware. The ghost in the machine is not as absurd as Ryle and Dennett want us to believe.

References

+ Francis Crick, Astonishing Hypothesis: The Scientific Search for the Soul, Scribner, 1994
+ Jay L. Garfield, Losing Ourselves: Learning to Live without a Self, Princeton University Press, 2022
+ Daniel Dennett, Consciousness Explained, Little, Brown and Co., 1991
+ György Buzsáki, The Brain from Inside Out, Oxford University Press, 2019
+ Gilbert Ryle, The concept of mind, University Of Chicago Press, 1949
+ Jean-Paul Sartre, Being and Nothingness, Routledge, 1943
+ John Searle, Minds, Brain and Science, Harvard University Press, 1984
+ Douglas Hofstadter, I Am a Strange Loop, Basic Books, 2007
+ Steven H. Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, Westview Press, 2014
+ Alan Garfinkel, Jane Shevtsov, Yina Guo, Modeling Life – The Mathematics of Biological Systems, Springer, 2017
+ Steven Rose (Ed.), From Brains to Consciousness, Princeton Univ. Press, 1998
+ Colin McGinn, The Problem of Consciousness, Blackwell, 1990
+ Susan Blackmore, Consciousness: An Introduction, Oxford University Press, 2003
+ Sidney Shoemaker, "Self-Reference and Self-Awareness." Journal of Philosophy 65 (1968) 555-567

The problem of free will - the agency paradox

Core points:

- Agency Paradox: if our choices are determined and caused by previous events, we are not in control; but if our choices are not caused by anything, they are random, and we are also not in control
- Free will must exist because it is the prize that marketing and advertising want to take away. Would PR, marketing and advertising all try to take away our free will if it does not exist?
- Consciousness is the tool to solve the problem of free will. First by a process named weighting reasons, and second by self-forming actions, by choosing what kind of person we want to become in the long run.

Free will is possible

> “Free will in the traditional sense is the power to be the ultimate creator and sustainer of one's own ends or purposes”
> ~ Robert Kane in “The Significance of Free Will” (Kane, 1996)

The ultimate paradox of free will is this: if our choices are determined and caused by previous events, we are not in control; but if our choices are not caused by anything, they are random, and we are also not in control. If we are really in charge of our own actions, then the actions can neither be fully determined by previous or historic events, nor can then be fully random.

A lot has been written about free will in the last centuries. Among the different standpoints and theories that emerged are compatibilism, semicompatibilism, libertarianism, hard incompatibilism, and revisionism (Fischer, 2007). And yet it has remained so far one of the most fiercely contested debates in all of philosophy and cognitive science (Kane, 2011). Some say free will is possible (List, 2019) others doubt it.

Meghan Griffith writes in her book about free will that we are beings who must constantly choose what to do next. Even if we refuse to make a decision this refusal would itself be a choice. We are "beings that must constantly choose" (Griffith, 2021). The question is do these choices leave us enough room for choices that are genuinely our own? Griffith claims science cannot tell us for sure whether we have free will or, if we do, what it is like.

In this chapter we would like to take a pragmatic look at the problem of free will. The three American founders of pragmatism William James (1842–1910), John Dewey (1859–1952) and Charles Peirce (1839–1914) agreed that free will is possible. They argued free will can be possible without stepping outside nature. It becomes possible by increasingly sophisticated forms of organization within nature. Dewey argues freedom consists in our capacity to intelligently direct action within the conditions of nature. We have “freedom in and among actual events, not apart from them” (Dewey, 1922).

Charles Peirce asks how increasingly sophisticated systems of interaction, habit formation and interpretation can evolve. One of his insights is the universe evolves from absolute randomness and chance into increasingly complex and organized patterns ruled by habits and law, driven by continuous interpretation and connection (Peirce, 1923). A simple organism like a reptile operates by hard-wired behavior patterns

Stimulus → Response

Peirce says the world contains spontaneous variation and continuous change. Chance leaves room for novelty. This room allows more sophisticated organisms to learn new habits. A more sophisticated organism like a mammal operates by learned patterns. Piece calls this “method of tenacity” (Peirce, 1923):

Stimulus
→ habit
→ response

Selecting the right one among multiple possibilities require effort and attention. An organism can either pick the one that feels right (Peirce's “a priori method”) or the one that society recommends (Peirce's “method of authority”), see (Peirce, 1923).

In both cases selecting the right one requires sustained effort, as William James writes. William James defines free will as the subjective capacity to voluntarily sustain attention on a mental object: “"The essential achievement of the will, in short, when it is most 'voluntary,' is to ATTEND to a difficult object and hold it fast before the mind [...] Effort of attention is thus the essential phenomenon of will” (James, 1918).

A conscious organism according to James and Peirce operates like this: it requires James's sustained effort, and fits to Peirce's "method of authority" or his "a priori method":

Stimulus / problem
→ representation of possibilities
→ selection
→ response

An intelligent organism has agency and operates like an agent in a continuous perceive-reason-action cycle. It creates hypotheses, evaluates them, and does it again in a loop until the best one is selected. John Dewey defines this intelligent deliberation as a "rehearsal" where an action is "tried out in imagination, not in overt fact" so that thought "foresees outcomes and thereby avoids having to await the instruction of actual failure and disaster" (Dewey, 1922).

Such an intelligent organism requires James's “sustained effort”, Dewey's “intelligent deliberation”, and is like Peirce's "method of science" which is basically the scientific method of formulating and testing hypotheses against the real world:

Stimulus / problem
→ hypotheses
→ anticipated consequences
→ experiment
→ revised hypothesis
→ response

Basically, having free will is the freedom and the capacity to acquire, modify and break habits. Free will can be seen as the ability of a system to modify the very mechanisms that determine its future behavior. No free will means being trapped in rigid habits and stubborn patterns of behavior or having no options to break these habits. Free will requires self-modification, i.e. free will requires the capacity of a system to modify its own behavior. A robot can have free will too if it can set its own directive. Humans can have it because we can modify our intentions. We can break the patterns that govern our behavior, at least in principle.

In other words, free will means the ability to break the relentless stubbornness in doing - or continue this way of doing if it is your second order volition, i.e. if it is what you want to want in Harry Frankfurt's words (Frankfurt, 1998). This does not mean achieving free will is easy, because breaking the patterns that govern the own behavior is hard. We will take a look how achieving free will is possible in detail in the next chapters.

Free will as an achievement

Free will means that we cause our own actions and make our own decisions. But who is really in charge of your actions and decisions? Is it yourself or someone else? The answer is of course it depends. Often it is yourself who is in charge, but yourself in the form of your own body, i.e. your preferences, emotions, desires and drives. Sometimes it is the opposite, and we are subject to others or random events. And then there are rare moments in between where we able to achieve free will. In any case the own self seems to be at the core of the irreducible agency paradox of determinism and randomness.

Schopenhauer said “Man can indeed do what he wants, but he cannot want what he wants" (in the German original “Der Mensch kann zwar tun, was er will, aber er kann nicht wollen, was er will”). Schopenhauer's famous quote means in a free world you are free to follow your desires, but you are not free to choose what those desires are in the first place. You have no control over where those initial desires came from, and you cannot consciously choose or change your underlying wants, cravings, preferences, desires or drives directly. They are determined by your biology, biochemistry, personality, psychology, history and upbringing. The underlying wants, cravings, preferences, desires or drives are connected to who you are, to your character and personality. Together they represent what you want. They are your will so to speak.

A "will" is never completely free because it is always someone's will. Either your will, your father's will, or god's will - the Lord's Prayer contains the line "your will be done, on earth as in heaven". This means a will belongs to a specific subject that wants something. It is basically what a subject wants. What a person wants cannot be directly changed by the subject itself, as we have argued in the last paragraph. It is attached to the identity, character and personality of a person, to the individual preferences, desires and drives. The moment a will becomes detached from your specific identity so that it can be "free," it ceases to be your will. Your will is inextricably tied to who you are.

From this point of view the idea of free will is almost an oxymoron, a contradiction. If your choices aren't determined by who you are - your values, memories, brain chemistry, emotions, personality and character - then they are not yours. It is not your will. But if your choices are strictly determined by who you are, by your personal psychology, biology, and past experiences, then your actions are not completely free.

By default, every animal looks constantly for food and mates. Whenever something is happening in nature it is either dinner time or pairing time. Emotions are the natural built in biological mechanisms to get food and mates when we need it (to maintain our bodies and to reproduce ourselves). Imagine we are able to suppress our emotions to obtain freedom. If we manage to suppress our emotions, we have the freedom to act without emotional influences, but we also lose the very mechanism that tells us what to do and gives our life a direction.

It is like a puppet that cuts its strings. It is free of the strings from the puppeteer but also has no longer someone that shows the direction in which to move. Another comparison would be an astronaut who uses a rocket to leave Earth and travels to the moon. He is no longer affected by gravity, but there is also nobody who gives him a direction where to go. In this sense asking for free will is like asking for a top or a bottom in zero gravity.

Nevertheless, free will is still possible. If we stay in the rocket picture, we can say it works like this: you fly into the orbit but do not leave Earth. Looking back, you see the world, including yourself. You see where you started, where you are, and where you are headed. You pick a spot which will make a difference for you if you land there. You act self-aware on the environment, and the environment acts back on you.

In practice a person is not using a rocket but just ordinary language which allows us to think and to record thoughts in books. A book can transport the reader in a different world, as Carl Sagan wrote about in his book Cosmos:
"one glance at it and you're inside the mind of another person, maybe somebody dead for thousands of years. Across the millennia, an author is speaking clearly and silently inside your head, directly to you".

Thomas Nagel argued in his book “View from Nowhere” that free-will seems to be only present in the subjective, first-person view and absent in the objective third-person view (Nagel, 1986):

- Subjectively, from the first-person view, we feel like we make choices and we are responsible for our actions
- Objectively, from the third-person view, we see ourselves as mere links in a chain of cause and effect, driven by upbringing, and the laws of physics, biology, and neuroscience

Do we have free will or not? If we consider that our behavior is determined by our upbringing, our past and our current environment and we are hardwired to act as our hardware (the character and personality encoded in our neural networks) demands, then the answer would be no. Neuroscientists know we are biologically biased to repeat past behaviors and rely on unconscious programming simply to save energy. This makes our daily decisions predictable, especially to those who know us well (Critchlow, 2019).

If we consider that we have consciousness and can change then the answer would be yes. It is a yes in the indirect sense that we have the freedom to choose what kind of person we want to become - at least in principle. It is an illusion to believe we can act "against our hardware". What kind of person we are decides what we do. In a sense we are all addicts who look for their drug, only the addiction differs. But what we do also determines what kind of persons we become. Free will is possible, but achieving it can be an art. An achievement.

Free will is tricky because we try to find it in our everyday, easy decisions, although this is not the place where to find it. According to Kane most of our daily actions are determined by our existing character, motives, and purposes. However, Kane says we are still free because we are ultimately responsible for having formed that character in the first place, and free will matters can be found when you are profoundly torn between two paths (Kane, 2024).

It is also tricky because directing others can be easier than directing ourselves. Solomon's paradox for example is the observation that people can give reasonable advice to others but fail to apply this advice to their own situation (Kross, 2021).

Some scientists argue it is not only tricky, but also hopeless. Robert Sapolsky (Sapolsky, 2023), Sam Harris (Harris, 2012) and Daniel Wegner (Wegner, 2002) claim that free will is an illusion and a biological impossibility.

Others like Alfred R. Mele say says science has not proven free will is an illusion. Conscious considerations genuinely influence and cause physical actions (Mele, 2009). Mele says free will just requires our conscious intentions to play a causal role in what we do. He argues that being influenced by our brains or our surroundings does not mean we are helplessly controlled by them (Mele, 2014). Even if the physical universe is entirely deterministic, we still possess according to Mele a practical, robust form of free will based on conscious, uncoerced deliberation. If we consciously weigh the pros and cons of a decision, we are collecting what Mele calls "motivating reasons" - considerations that speak in favor of doing something or against it. Conscious spotlight can help to recognize the motivating reasons more clearly, and if you select a best choice in this light without being forced, you are exercising a form of free will. Rational deliberation can influence our choice, and this would be the free will that matters (Mele, 2022). This fits to the idea of Robert Nozick who claimed when we deliberate, we weigh various reasons in the spotlight of consciousness and assign a weight to them. The weight of the reasons then determines our choice and the decision we make. (Nozick, 1981) The solution to the problem of free will sketched in this chapter follows roughly the original ideas Robert Nozick, Robert Kane, and Alfred R. Mele.

Daniel Dennett argues in his book “Freedom Evolves” (Dennett, 2003) that even in a deterministic universe we have the free will that matters: the free will to make decisions without duress. He uses free here in the common-sense of free of being forced to do something: we have free will when no outer force compels our actions.

Why we have this free will at all is an interesting, difficult question. On the lowest molecular and cellular level actions and decisions are deterministic, which could lead to the conclusion that free will is an illusion. And yet most of us would agree with Kevin Mitchell that “we are not mere machines responding to physical forces but agents acting with purpose” (Mitchell, 2023).

When does deterministic behavior turn into unpredictable, indeterministic behavior? The basic question in every moment for a living being is: where does it want to go? What should it do next? There is no easy answer to this complex question. How we decide what to do depends on a few factors. Our decisions and choices are influenced by internal and external factors and by the wiring of neural circuits in our heads. Internal factors include the chemistry of the brain and the current mood. The large number of possible combinations of internal and external factors that influence our decisions leads to a combinatorial explosion of possible actions which offers room for the "free will" of many different personalities and distinct characters. The free will of a living being consists in the possibility to choose a specific action from a range of different options. The choice reflects the will – what the living being wants to do at that moment. A living being will primarily make decisions that correspond to its own intentions and will lead in the current situation to the best results based on personal experience. Even if we have the free will to make decisions without duress, we cannot change the wiring in our heads immediately (which reflects our character and personality, our likes and dislikes, and our general preferences). We can try to change our habits, and the habits in turn are able to change the wiring. We can also try to learn new things which change our perspective and influence our decisions. So, it is fair to say that the situation is complicated.

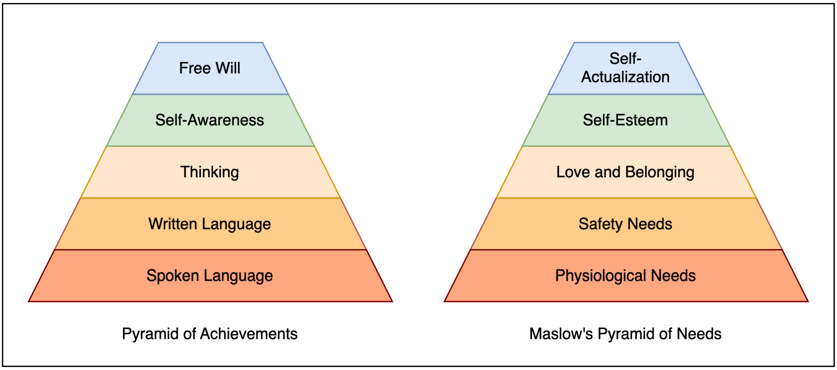
Maybe it is more helpful to approach the problem from the other side and ask when we are free of external influences and how we can avoid acting under duress? Free will means people could have done otherwise and could have acted differently – if they would have made different decisions in the same circumstances. It is primarily language and consciousness that enable us to act differently. Our language gives us the freedom of choice to act in one way or another, because it allows us to think about different options and outcomes, and it allows us to realize that others are trying to influence us. Daniel explains that imagining and reflecting on the outcomes of possible actions are the key here. In his autobiography "I've been thinking" he says

> "Free will […] is an achievement, not a metaphysical feature, of normal human beings who have learned how to control their many degrees of freedom – and to keep others from remotely controlling them – be developing such habits of self-stimulation as imagining and reflecting on the outcomes of possible actions."
> ~ Daniel Dennett in "I've been thinking" (Dennett, 2023)

Daniel describes free will as an achievement. This pinnacle of achievement is only possible if humans have mastered other, more mundane achievements: if they are able to speak, to read, to write, to think and finally to be self-aware, which is an achievement too. We can think of it as a pyramid of achievements which is very similar to Maslow's pyramid of needs

Maslow proposed that human beings are motivated by unsatisfied needs, which must be met in a general order from most basic to most advanced. While popular culture always displays this as a rigid pyramid, Maslow described it as a fluid, overlapping system where multiple needs can be partially met at the same time. However the general order remains the same: the most basic needs must in general be met first before the more advanced needs are considered (Maslow, 1954).

For the pyramid of "achievements" it is similar. The most basic achievements must be completed before the more advanced ones can be tackled. Free will is not possible without self-awareness, self-awareness is not possible without independent thinking, independent thinking is not possible without language, etc. The ultimate achievement in this view is free will.



Free will is indeed a valuable asset, not only because it is the ultimate achievement, but also because its manipulation has power. If it is manipulated then people can be influenced to do what others want: join a group, vote for a party, or buy a product. There are all kinds of psychological tricks that can be used to influence our decisions. One psychological trick mentioned by Robert B. Cialdini (Cialdini, 2021) is for example reciprocity: we feel obligated to return favors, even uninvited ones. For instance, we are much more likely to buy a Salami from a salesperson who offers us a free slice of Salami to taste, because we feel we need to return the favor.

This means a salesperson that offers a small, unexpected gift can influence customer to buy his products.

On a larger scale this manipulation is invaluable for religious groups, political parties and economic companies. Free will is the treasure island that everybody wants to own. Free will is the treasure that everybody wants to take away. Religious groups want to manipulate our free will to do whatever they preach. Parties want to manipulate our free will to vote for them by PR firms and ads. Companies want to manipulate our free will to buy their products. These industries have always walked a fine line between "acceptable persuasion" and "manipulation" (Sunstein, 2025).

Advertising and marketing industry

There is in fact a whole 100-year-old industry which has evolved to manipulate our free will: since Edward Bernays wrote his book about propaganda and PR in 1928, the emerging PR/advertising/marketing industry tries to influence our decisions in order to do what the advertisers want: follow their movements, join their groups, vote for their parties, buy their products, etc. This is like the hard problem of consciousness. There is a whole 100-year industry which has evolved to solve the hard problem of consciousness: the film industry and all forms of show business which earn money by showing us what it is like to be someone.

It turned out that the solution to the biggest questions in philosophy can not only be found in the most ordinary things, but the big problems of philosophy have led to giant industries over the last century. Show business grants us a glimpse of what it is like to be someone else (whether Indiana Jones hero or Han Solo rebel) and the advertising industry constantly tries to take away our free will. The advertising industry has built large free-will-manipulation machines that manipulate our opinions in subtle and sophisticated ways by showing us ads in all kinds of channels and all kinds of devices.

The best argument that we have free will is the existence of a whole industry that tries to influence it, isn't it? Therefore, we can say free will must exist because it is the thing that marketing and advertising want to influence, the prize that they want to take away, the prey they want to catch.

Why is it so hard and one of the fundamental problems? Because there are multiple problems we need to solve and insurmountable obstacles that impede us if we want to find our own way to go. To achieve free will we need to be

- free of basic needs that distract us (e.g. the bottom of Maslow's Hierarchy of Needs)
- free of hidden influences by external forces & from goals imposed on us
- free from illusions and self-deceptions

If all these conditions are fulfilled, we need to be

- able to formulate our own intentions (which requires language)
- able to perceive ourselves in the third person view (which requires self-consciousness)
- able to set our own directive (which is tricky)

We have a chance of free will only if we can perceive ourself and the forces acting on us in the third person view. This allows us in principle to set our own directive. But this is tricky because as Schopenhauer noticed "A man can do what he wants, but not want what he wants" ("Der Mensch kann tun, was er will, aber er kann nicht wollen, was er will"). All these obstacles make you wonder why free will is possible at all. And yet we have the ability to break the patterns that govern our behavior. We have the freedom to choose what we want to be on fire about. The question is how?

To be able to think about achieving our own goals and intentions in the first place we must be free for a moment from the constant needs and desires of our biological bodies. If the bottom of Maslow's hierarchy of needs is not fulfilled than there is no room for free will left. A homeless person in San Francisco thinks only where he can sleep and what he can eat, while a billionaire can do whatever he wants. He can even use champagner for the shower on his superyacht, as Gregory Salle describes in his book "Superyachts: Luxury, Tranquility and Ecocide".

One obstacle is to be free of hidden influences by external forces and from goals imposed on us. What are the hidden forces which try to influence our decisions (thereby reducing our free will) and how can we resist? As described in the last article, advertising and marketing play an important role here. They all want to influence our free will and impose goals on us we are supposed to want (like you need to eat that burger or buy that t-shirt). How this subtle persuasion is done is explained in countless books like “Positioning: The Battle for Your Mind” or “Influence: The Psychology of Persuasion” or “Advertising, The Uneasy Persuasion” and many others. Thus, the question should be “who has free will?” Obviously the rich and those who are smart enough to resist manipulation by marketing, advertising and propaganda have much more free will than the rest.

Another closely related obstacle is to get rid of the illusion that we know what we want. Erich Fromm says in his book “Escape from Freedom” (Fromm, 1941) that “Modern man lives under the illusion that he knows ‘what he wants,’ while he actually wants what he is supposed to want. In order to accept this it is necessary to realize that to know what one really wants is not comparatively easy, as most people think, but one of the most difficult problems any human being has to solve. It is a task we frantically try to avoid by accepting ready-made goals as though they were our own.”

Then there is the obstacle to perceive ourselves from the third-person view, which is of course difficult, because it leads to self-consciousness, and this is a source of confusion. As Gilbert Ryle says, knowing yourself is tricky. He writes “Should I, or should I not, put my knowing self down on my list of the sorts of things that I can have knowledge of? If I say ‘no’, it seems to reduce my knowing self to a theoretically infertile mystery, yet if I say ‘yes’, it seems to reduce the fishing-net to one of the fishes which it itself catches” (Ryle, 1949). In his book Gilbert Ryle positions himself as a “ghost buster” by arguing there would be no ghost in the machine. Ergo no ghost in the machine which could have a free will either. The difficulty is despite all this confusion to develop the ability to think about yourself in order to define your own intentions.

Setting your own intentions and goals, your own directives is the ultimate form of free will, isn’t it? William James says in “The Principles of Psychology” in Volume 2, Chapter XXVI “WILL”: “The essential achievement of the will, in short, when it is most ‘voluntary,’ is to ATTEND to a difficult object and hold it fast before the mind”. He continues “though the spontaneous drift of thought is all the other way, the attention must be kept strained on that one object until at last it grows, so as to maintain itself before the mind with ease. This strain of attention is the fundamental act of will.” (James, 1918).

He suggests that the act of will is fundamentally an effort of attention because minds wander and thoughts drift, like particles who are subject to constant random movement. Our ability to focus attention, especially in the face of competing distractions, is therefore central to our experience of exercising free will. Free will is according to James related to the question “what are the forces that increase or influence our attention”. Is it us – by setting our own goals and our own directives – or something else?

Since we cannot change our own intentions, desires and directives directly, we usually need to do it indirectly by interacting with someone or something which acts as an intermediary. It can be as simple as a calendar entry or a sticky note which tells our future self in the next few days what to do. Or it can be praying (collectively or individually) which is essentially talking to our future selves, as mentioned in the previous book I have written. Talking to our future selves helps to set our own directive. If we remember what we said or thought in our prayer then our past selves tell us what to do and what to avoid. In this sense it helps to realize the will incorporated in the remembered text, usually in religious contexts the will of an higher entity, but it can also help to realize the will of ourself, and thereby help to enable free will, volition and willpower.

Just how does it work? Let us take a look at the philosophical explanations of two American philosophers, Robert Nozick (1938-2002) and Harry Frankfurt (1929-2023).

Consciously shaping our own character

Harry Frankfurt was a philosopher at Princeton University, Robert Nozick a philosopher at Harvard University. Like Daniel Dennett and Gilbert Ryle he tackled the big, fundamental questions using ordinary language philosophy, for example in his book “Philosophical Explanations” (Nozick, 1981). It has a chapter about “Free Will” which Daniel liked. Let us have a look at it.

Robert Nozick argues that if we really have free will our actions and decisions can not completely be determined by external forces, and cannot be completely random. They must be determined by ourselves, by internal forces, by playing an active role and making decisions that define our future behavior.

In his book “Philosophical Explanations” (Nozick, 1981) he argued a free decision by an agent can be made by consciously thinking about which choice is best. Before deciding, an agent weighs the different reasons, which means comparing pros and cons associated with their available choices. The agent then decides consciously how much weight a particular reason holds and assigns weights to it. The act of making a choice is the act of dynamically assigning more weight to one set of reasons over another. After weights have been assigned the best choice emerges automatically.

Modern neuroscience says this weighting of reasons is real and happens in the ventromedial prefrontal cortex (vmPFC). Emily Falk argues whenever we face a decision our brain runs a real-time, often subconscious calculation in the vmPFC to determine which option has the highest net value. She explains that we can alter our decisions by changing the variables that go into our brain's calculation: by expanding awareness and changing perspectives we force the vmPFC to recalculate the subjective value of our choices (Falk, 2025). In other words, conscious reasoning acts like a spotlight that enables us to recalculate the best choice in a complex environment.

It is possible to model his concept of "weighing reasons" mathematically, either by using Quantum Cognition where the act of deciding equals a wave function collapse (Busemeyer, et al., 2012) or by a Neural Network Attractor Model based on a Wilson-Cowan model where the firing rates of wo mutually inhibiting pools of neurons represent two different choices. The Wilson-Cowan model is a pair of coupled differential equations that describes the interaction dynamics between populations of very simple excitatory and inhibitory model neurons (Dayan, et al., 2001) .

In the context of modeling free will, the Quantum Model and the Neural Network Attractor Model can be viewed as two different descriptions of the same underlying cognitive process, viewed at different scales, like Schrödinger's wave equation and Heisenberg's matrix mechanics which were later to be shown identical.

We can in fact use a similar model as we have for insights in humor in the earlier chapter. The third assembly which is responsible for the sudden shift in perspective corresponds here to the conscious spotlight of conscious deliberation assigning additional weight to A or B

Component	Insights in Humor	Decision in Free Will
Assembly A	Setup of interpretation 1, part A of conventional meaning	Choice A
Assembly B	Setup of interpretation 1, part B the conventional meaning	Choice B
Assembly C	The hidden, alternative meaning of interpretation 2	conscious deliberation assigning additional weight
Tipping/Decision Point	Punchline triggered by C, "Aha!" moment	Moment of choice triggered by C
Result	Paradox resolved (Laughter)	Choice made (Action)

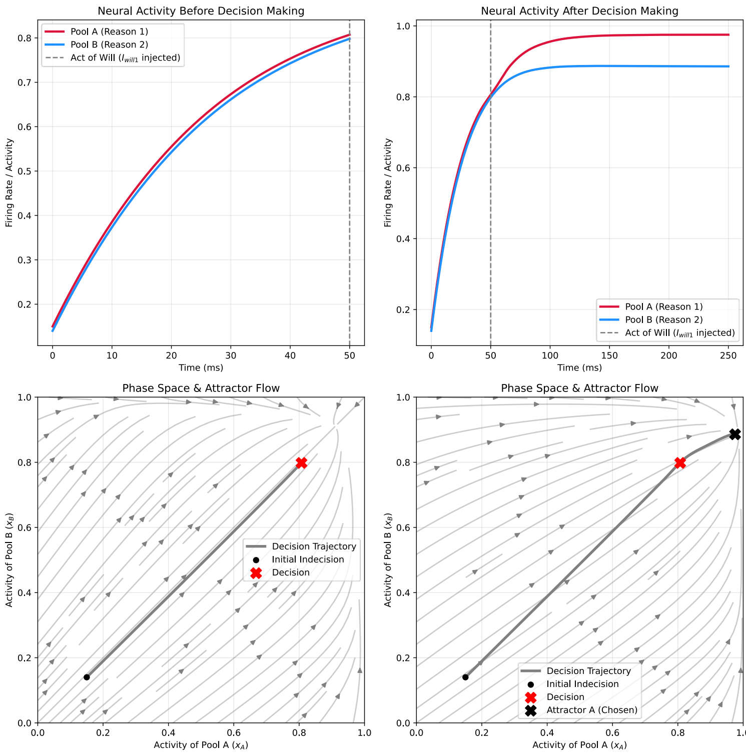
The difference is in humor we have a stronger inhibition between A and B because of the incongruity that needs to be resolved, and the shift is triggered by an external impulse because the punchline is triggered by external input:

External Punchline
→ Triggers C
→ Warps Landscape A/B
→ Paradox Resolved (Laughter)

In free will as Nozick and Falk describe it we have a proactive shift triggered by internal consideration and the spotlight of conscious deliberation:

Internal Deliberation
→ Activates C
→ Warps Landscape A/B
→ Choice Made (Action)

The dynamics looks like this (see Appendix B): up to a decision point both pools have the same average rate of activity. At the decision point the reasons for one pool or choice are getting more weight, i.e. this pool gets a boost. This boost changes the whole dynamic, and the system moves toward the attractor where one pool wins.



This kind of conscious decision making is complicated and requires thinking, for example imagination of alternatives and consequences. But in the end weights are assigned not by conscious thinking but by calculating how much value and valence each choice has for us, which is the function of emotions. Even if conscious reasoning helps to become aware of all consequences our emotions make in the end the decision for us, and not our self-aware self.

Decision making for ourselves can be made in simpler ways too, for example indirectly over the environment. We act on the environment – for instance we are saying a prayer or writing a note or creating a calendar entry which reminds us to do something, or signing a contract which forces us to do something - and then the environment acts on us in the future - if we remember the prayer, read the note, stumble on the calendar entry, or are reminded of the contract which we have signed. This detour over the environment allows us to solve the paradox that an action must be undetermined and yet not random, even if there is no mysterious downward causation, because the causal chain runs from the self over the environment back to the own self.

For more complex behavior it works in a similar way. Nozick argues that in each moment we try to detect the best possible action in the current situation, given the hierarchy of desires, principles, and intentions which reflect our personality and our character. He says “we always do what we most prefer” and calls this “tracking bestness” (Nozick, 1981), which means to detect the best choice of action giving a system of values. This fits again perfectly to what neuroscientist Emily Falk has found out later (Falk, 2025). One could say Nozick was ahead of his time.

What we perceive as good and bad depends of course on our emotions, but also on the social and moral systems that have shaped us, i.e. by our character. The character of a person is defined by a hierarchy of desires, directives, principles, and intentions which reflect the personality and the character of this very person. An altruistic and pious person for instance will value moral laws higher than a selfish and vicious person. What you care about – this could be beauty for an artist, honesty for a scientist, accuracy for a perfectionist, or fairness for an altruist – is essentially your self, encoded in a hierarchy of volitions. The self is a complex structure that channels causal chains in a way that reflects the person’s identity. Nozick says “Our principles fix what our life stands for, our aims create the light our life is bathed in, and our rationality, both individual and coordinate, defines and symbolizes the distance we have come from mere animality” (Nozick, 1993).

Free will according to Nozick is possible because there is a connection between the agent’s self and the action. The key here is the process of reflecting, endorsing, and identifying the desires and principles which guide our own actions. By agreeing with the higher-order directives and principles that control the causal chain of action an agent can identify itself with the source of the acts. Nozick calls this “acts in equilibrium”. If an agent does not endorse the hierarchy it can refuse to act. Nozick describes this situation as “acts in disequilibrium”.

Harry Frankfurt calls such a refusal to act a second-order volition against that first-order desire (Frankfurt, 1998). He argues we do not just have desires, but care about which desires move us. What makes us persons is that we do not just have desires, but we are able to care about which desires move us. He calls it a second-order volition, and describes a second-order volition as a desire that a certain desire be one's will. He regards nonhuman animals and very young children without second-order volitions as agents which follow their instincts and desires blindly (either without being aware of them or without questioning them), and calls them "wantons".

For Nozick, freedom of the will is basically the freedom to choose what kind of person we want to become and which way we want to go in our life. We can shape our own character and form our personality gradually through our actions and choices, and this self-forming capacity for conscious self-formation is what makes us morally responsible and eventually free. The more our personality grows in a certain direction, the more it shapes our actions in this way.

This makes only sense if we can choose what person we want to be. At least in principle it is possible, because we can certainly more or less influence who we become, but of course it can be difficult in practice to change our own character for all the reasons mentioned before and because self-control is not an easy problem. Nozick calls the process of choosing a person we want to be "self-choosing". He says "the fullest autonomy is had only by a being whose essence is self-choosing".

Robert Kane calls the essential parts of this process "self-forming actions" in his book "The Significance of Free Will" (Kane, 1996; Kane, et al., 2022). In his later book "Do we have free will? A debate" (Kane, et al., 2022) he defines self-forming actions as those "acts by which we form and re-form our wills (our characters, motives and purposes) and for which we could have done otherwise, which must occur at some times in our lives, if we are to be ultimately responsible for having the wills we have and hence for being the kinds of persons we become".

"Self-forming actions" (SFA) help to form the character of the agent and help to set new directives or intentions. Kane writes "SFAs must be 'will-setting'". Both Nozick and Kane agree that even if our actions are determined by our character, these actions can be free to a certain extent if we are able to shape and form our own character. These "self-made selves" (in the words of Daniel Dennett) explain how we can be moved by "causal chains of events" and yet have the freedom of choice.

An important point is that a character and a personality cannot be changed in an instant. The self-forming part takes much longer, while the decision in a self-forming action can be short and depends on conscious moments of self-awareness. In these short moments we can start something (like signing up for a course or at a school, or joining a club or company, etc.) which helps to form our future self. In the moment of the self-forming action itself we see indeterministic free-will in action, and once the self is formed the action is deterministic.

Nozick says what matters is how the self stands in relation to the causal chains that control our action. When the agent reflects, endorses, and identifies with the desires and principles guiding the action, and it “acts in equilibrium”, then there is a close connection between the self and the action. Harry Frankfurt says this equilibrium is reached when your first-order desires align with your second-order volitions – i.e. when you want what you want to want – and then we are authentically and wholeheartedly ourselves. He argues what we care about is important: we have free will because we care about our will. Frankfurt says “A person’s free will is free only if he is free to have the will he wants” (Frankfurt, 1998). Nozick would agree.

Moral emotions

English philosopher P.F. Strawson has shifted the debate about free will towards a debate about moral responsibility. He has argued in his essay “Freedom and Resentment” (Strawson, 2008) that moral responsibility is a basic practice of our social life where we hold each other responsible, based on our social emotions like resentment, anger, pride, shame, praise, and blame.

Basic emotions like fear (helps us to avoid dangerous situations), disgust (helps us to get rid of unhealthy food), and hunger (helps us to search for food) help to maintain the integrity of our body. Bad feelings are, as discussed earlier, adaptations to bad situations and thereby good for us. Social or moral emotions like sorrow, sadness, resentment, anger, outrage, pride, shame, remorse, guilt or gratitude help to maintain the integrity of our social fabric, i.e. they are essential for the fabric of human relationships.

Oatley et al. (Oatley, et al., 2020) argue that self-conscious emotions preserve group membership. Shame and guilt signal remorse when an individual breaks a group rule, preventing expulsion. Pride signals high status and expertise, while embarrassment functions as a nonverbal apology to appease offended group members. Gratitude functions as a social glue and increases long-term group cohesion.

Moral emotions arise through social interactions. Adam Smith wrote about moral emotions in 1759, Charles Darwin in 1871, and recently Robert Frank (Frank, 1988) and Jonathan Haidt (Haidt, 2003). Adam Smith saw gratitude and resentment as important moral emotions because they help to maintain social order. They are guardrails that keep us on the right track of socially acceptable behavior.

Darwin saw moral emotions as evolved traits, shaped by natural selection to help humans live, cooperate, and thrive in complex social environments. He writes “the first foundation or origin of the moral sense lies in the social instincts”. Later they were supplemented by a personal “conscience” which became “the supreme judge and monitor” of social behavior, causing shame, remorse, regret and guilt. He says regret and shame are caused by the consciousness of wrong behaviour and writes in his book “The Descent of Man”:

> At the moment of action, man will no doubt be apt to follow the stronger impulse; and though this may occasionally prompt him to the noblest deeds, it will more commonly lead him to gratify his own desires at the expense of other men. But after their gratification when past and weaker impressions are judged by the ever-enduring social instinct, and by his deep regard for the good opinion of his fellows, retribution will surely come. He will then feel remorse, repentance, regret, or shame; this latter feeling, however, relates almost exclusively to the judgment of others. He will consequently resolve more or less firmly to act differently for the future; and this is conscience; for conscience looks backwards, and serves as a guide for the future.
> ~ Charles Darwin, The Descent of Man, 1871

In chapter 7 “The evolution of moral agency” of his book “Freedom Evolves” (Dennett, 2003) Daniel Dennett draws heavily on the work from Robert Frank, and quotes for instance this paragraph which says that “Moral sentiments may be viewed as a crude attempt to fine-tune the reward mechanism, to make it more sensitive to distant rewards and penalties in selected instances". Jonathan Haidt goes a step further and argues that moral emotions are not only fine-tuning of an existing reward mechanism, they are a reward mechanism for a new system, because they are about "the interests or welfare either of society as a whole or at least of persons other than the judge or agent" (Haidt, 2003).

In this new social system which Haidt mentioned, the freedom of the agents is "both their greatest strength and their greatest problem", to use the words of Dennett. It is their greatest strength because they can be much stronger if they cooperate, and their greatest problem is that they can decide to cheat and deceive. What is needed is therefore some kind of social alert if someone is cheating or deceiving others – which is one function of the moral emotions.

And here we come back to Peter Frederich Strawson who says moral emotions are essential since they regulate our relationships and shape how we hold each other responsible (Strawson, 2008). He calls them "reactive attitudes" because they cause us to react in certain situations in specific ways. Others expect us to act in certain ways depending on our relationship. We are morally responsible if we meet the expectations of others or break them. For example, family members and friends are supposed to help each other. If a friend promises me to help because I move to a new apartment, and on the day of the move nobody shows up, then I will probably feel resentment against him, because he broke his promise, and he will feel guilty, because he promised to help but did not.

As Daniel Dennett argues in his two books on free will "Elbow Room" (Dennet, 1984) and "Freedom Evolves" (Dennett, 2003), we have free will because we are free in all the ways that matter, even if our actions might be determined to a certain degree or caused by other events. Because we are free we can keep or break promises. We can try to be friends or not. We can help or refuse to help. And we hold individuals, especially our family and friends, accountable for their actions because they possess the cognitive machinery to understand the implications of their choices and because they could have acted differently had they chosen to. Social emotions help us to do that.

The fact that these moral emotions evolved at all – especially emotions like remorse, repentance, regret, or shame which Darwin described so well – implies we have the practical freedom to choose what we want to do (or at least who we want to become). They help us to hold others accountable which does not make sense if nobody is accountable because everybody has no choice. Let us say we have no free will. Then the social emotions would not work or would not make sense. They make sense because we think we – and others – could have acted differently. Consider for example the case of the friend who refuses to help. He feels guilty because he knows he could have acted better. I feel resentment or anger because I know he could have acted better. If there is no free will then these feelings make no sense, and we have a contradiction, because we know they exist.

Therefore it can be considered as evidence of free will. In other words the very existence of social and moral emotions is strong evidence of free will. Maybe even an indirect proof? We have argued the existence of a whole industry (PR, marketing and advertising) that tries to influence our free will is compelling evidence for its existence. The same can be said of moral emotions who influence our free will from the other, biological side. The very existence of moral emotions can be considered as evidence of free will. They would not have evolved if we would not have free will which allows us to form new systems. Systems for which we need new control and reward mechanisms. Viewed differently we could also say that moral emotions are an adaptation to a world where people have free will – including the will to cheat, to deceive or to defraud others. Strawson says "Our practices do not merely exploit our natures, they express them"]. They express that we see ourselves as agents who hold each other accountable because we have the free will that matters – to decide in each moment if we want to follow our selfish interests or want to respect moral obligations.

Why free will remains hard

Achieving free will sounds easy because if it just deciding to do something without being forced or manipulated. Our conscious, uncoerced, reasoned choices are in principle sufficient for free will, as Alfred Mele writes, because we have the freedom to choose Option A or Option B under the exact same circumstances (Mele, 2022).

But it remains hard because as we have seen

- many powerful actors are trying to take free-will away from us by manipulating us

- it is hard to make consciously a decision that contradicts what we are used to do and what we want to do

Roy Baumeister wrote that free-will is difficult because we are built to rely on automatic processes, habits, and immediate emotional responses, and we have strong innate drives geared toward survival and immediate gratification. He argues that free-will relies on self-regulation - the capacity to actively override your first impulse, but if you spend too much mental energy resisting temptations or managing emotions, you could experience mental exhaustion, decision fatigue or "ego depletion", and this would make it progressively harder to exercise free will. Free will is hard because staying deliberately focused on specific long-term goals requires a fragile balancing act between what we have to do and what we want to do (Baumeister, 2024).

We are usually hardwired to act as our hardware (the character and personality encoded in our neural networks) demands. And yet even if free-will is difficult to achieve directly, it is possible indirectly. We can try …

- to control ourselves indirectly over the environment (by a note, a calendar entry, a notebook, or anything that communicates with our future selves)
- to shape our own character by becoming a self-choosing person, as Nozick named it

We have the freedom to choose what kind of persons we want to become, at least in principle by self-forming actions, as Nozick and Kane say. While it is possible to choose the person we want to be and we can achieve autonomy by becoming a self-choosing character, it certainly requires a large amount of self-awareness, consciousness and discipline, and it is not easy to do.

A model of free will

> "Experience teaches us no less clearly than reason, that men believe themselves to be free, simply because they are conscious of their actions, and unconscious of the causes whereby those actions are determined"
> ~ Baruch Spinoza (1632-1677) in "Ethics"

Determinism and free are the ends of a dynamic spectrum as I learned from the FRIAM group members (Nick Thompson, Eric P. Charles, Steve Smith, Glen Ropella, Marcus Daniels, and Jon Zingale). This fits to the modern 21st century view of free will in fields like neuroscience (Haggard, et al., 2015) that human action can be seen as a continuous spectrum of conscious involvement, from pure, involuntary reflexes and habits at one end to conscious choices, voluntary actions and deliberate decisions at the other end.

The deterministic end of the spectrum is marked by low agency and activity in spinal cord and brainstem. The free will end of the spectrum is marked by high agency and activity in the prefrontal cortex (Haggard, et al., 2015).

The ultimate answer to the problem of free will is "it depends - how much you try to get it". If you do not think about what you want and kind of person you would like to be you are most likely blindly following your habits, biological needs and environmental triggers. If you do not actively try to obtain free will then you might believe you are free but as Spinoza said it is an illusion.

1. **The Unconscious State: a biological machine driven by selfish genes**

   If you don't think about it at all you only have an illusion of free will, just as Spinoza said. You are purely driven by habits, biological urges, and environmental triggers. Basically, the agent that defines our intentions can be found in the selfish genes as Richard Dawkins called them. They are the ones who influence our intentions by the emotions they construct.

2. **The Observational State: Watching the Machine**

   If you become aware of yourself you start to recognize that you don't have free will. Meta-cognition and self-awareness allow you to see you are NOT the cause of your own actions. It is rather your habits, your emotions and your biological urges that control you.

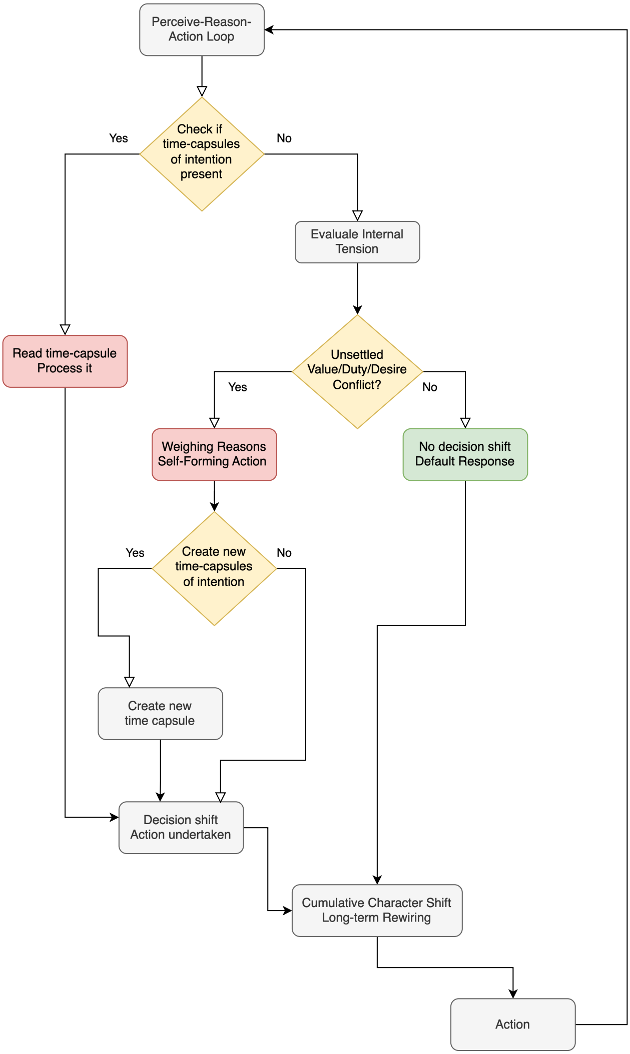
3. **The Interactive State: human agency by beating the machine and strategic intervention**

   If you have self-awareness how your system works, you can try to beat the machine in your head – by Robert Kane's SFAs and other tricks. While you cannot beat the machine completely, you can outsmart it from time to time by deliberate thinking and by time capsules of intention such as calendar entries, prayers and sticky notes.

These 3 levels or tiers correspond roughly to the classic theories of free will that are known in philosophy: determinism (tier 1), determinism & incompatibilism (tier 2), compatibilism (tier 3), see (Kane, 2005) or (Fischer, et al., 2007).

This model also solves the problem of agency. Like free will, agency can be seen as a spectrum of self-awareness and effort. The more we try to be the agent of our own actions, the more likely we are. Thus, the answer to the paradox of agency is a dynamic theory of agency. Some people have fewer cognitive degrees of freedom than others, and there are moments like Robert Kane's SFAs (self-forming actions) in action where we show full free will, and others where we completely lack it.

Achieving free will can be an art. It means to align a 3rd person action with a 1st person action. This requires consciousness, self-awareness and is harder than it sounds. It turns out that two of the hardest problem, the hard problem of consciousness and the hard problem of free will are related. Normally animals including ourselves perceive themselves in the first-person view and others in the third-person view. Social animals take a step further and perceive the own group in the first-person view and other groups in the third person view. The hard problem is to understand the first-person view in others (the hard problem of consciousness and subjective experience) and the third person view of ourselves (related to self-awareness and the problem of free will). From this point of view the two fundamental problems are related. Both involve switching from the subjective 1st-person to the objective 3rd-person perspective or vice versa. As Thomas Nagel said the difficulty is to reconcile the perspective of a particular person inside the world with an objective view of that same world (Nagel, 1986).


Perceive-Reason-Action Loop
Check if time-capsules of intention present
Yes
No
Evaluale Internal Tension
Read time-capsule Process it
Unsettled Value/Duty/Desire Conflict?
Yes
No
Weighing Reasons Self-Forming Action
No decision shift Default Response
Create new time-capsules of intention
Yes
No
Create new time capsule
Decision shift Action undertaken
Cumulative Character Shift Long-term Rewiring
Action

References

+ P.F. Strawson, “Freedom and Resentment and Other Essays”, Taylor & Francis, 2008
+ Pamela Hieronymi, Freedom, Resentment, and the Metaphysics of Morals, Princeton University Press, 2020
+ Adam Smith, The Theory of Moral Sentiments, 1759
+ Charles Darwin, The Descent of Man, 1871
+ Robert H. Frank, Passions within Reason – The Strategic Role of the Emotions, W.W. Norton & Company, 1988
+ Jonathan Haidt, The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), Handbook of affective sciences. Oxford University Press (2003) pp. 852-870
+ Daniel Dennet, Elbow Room - The Varieties of Free Will Worth Wanting, The MIT Press, 1984
+ Daniel Dennet, Freedom Evolves, Viking Books, 2003
+ Daniel Dennet, I’ve been thinking, W.W. Norton & Company, Inc., 2023
+ Robert Nozick, Philosophical Explanations, Harvard Univ. Press, 1981
+ Robert Nozick, The Nature of Rationality, Princeton Univ. Press, MIT Press, 1993
+ Robert Sapolsky, “Determined”, Penguin Press, 2023
+ Sam Harris, Free Will, Free Press, 2012
+ Robert Kane, The Significance of Free Will, Oxford Univ. Press, 1999
+ Robert Kane (Ed), The Oxford Handbook of Free Will, Oxford University Press, 2011
+ Kevin J. Michell, Free Agents: How Evolution gave us Free Will, Princeton University Press, 2023
+ Harry G. Frankfurt, The importance of what we care about: philosophical essays, Cambridge Univ. Press, 1998
+ Alfred R. Mele (Ed), Surrounding Self-Control, Oxford Univ. Press, 2020

The hard problem of consciousness – the paradox of subjective experience

Core points:

- The paradox of subjective experience: we understand subjective experience in ourselves. We can feel the pain of others, and yet we find it very hard to understand their subjective experience. And because it is hard we find it interesting.
- Cinemas and movie theaters are the tools we have invented to solve the hard problem of consciousness
- The hard problem of consciousness as the playground and business opportunity for show business and movie industry.

Can we solve the hard problem of consciousness, i.e. understand the subjective experience of others? No, not directly. Yes, we can. At least we can come close to solving it. Approximately. Here is why.

The hard problem

> "A wonderful fact to reflect upon, that every human creature is constituted to be that profound secret and mystery to every other. A solemn consideration, when I enter a great city by night, that every one of those darkly clustered houses encloses its own secret; that every room in every one of them encloses its own secret; that every beating heart in the hundreds of thousands of breasts there, is, in some of its imaginings, a secret to the heart nearest it! Something of the awfulness, even of Death itself, is referable to this."
> ~ Chapter 3, The Night Shadows of a "Tale of Two Cities" from Charles Dickens (1812 – 1870)

Already Charles Dickens pointed out in his "Tale of Two Cities" that human beings are a mystery to each other. Like every house which encloses a secret, every person "encloses" a secret: the personal universe of phenomenal consciousness and subjective experience.

One of the first who observed the hard problem – besides Charles Dickens - was Joseph Levine. Joseph Levine introduced the concept of an “explanatory gap” between 3rd-person descriptions and 1st-person subjective experiences. The explanatory gap exists according to Levine because physical descriptions cannot explain why or how a neural process produces a qualitative "feel". As a central metaphor for the disorientation, obscurity, and conceptual confusion that surrounds consciousness in the philosophy of mind he used in his book “Purple Haze: The Puzzle of Consciousness” the term "purple haze" which is named after Jimi Hendrix's iconic song of the same name (Levine, 2004). Jimi Hendrix's song is a love song about being completely confused and disoriented by infatuation. Levein writes

> “Consciousness presents us with a puzzle that seems to resist both reduction and non-reduction. We are caught in a purple haze: committed to materialism by our scientific outlook, yet unable to make sense of subjective feel within that materialist framework.”
> ~ Joseph Levine, Purple Haze: The Puzzle of Consciousness

In his later book "Quality and Content: Essays on Consciousness, Representation, and Modality" (Levine, 2018) he doubles down on the position that first-person views and third-person views are structurally incapable of being analytically integrated.

Levine is right. The hard problem is not completely solvable, and the explanatory gap that Levine described cannot be completely closed. But we can narrow it and achieve an approximate understanding, like the numeric solutions of atomic orbitals or numeric solutions to the three-body problem. Poincare showed that the three-body problem cannot be solved analytically using closed-form algebraic expressions or standard mathematical functions. But we can create approximate solutions by numerical calculations. We can come to an approximate understanding of the what-it-its-like-to-be question, an asymptotic approximation but not a complete analytical solution.

The philosopher that is known for coining the term “the hard problem of consciousness” is David Chalmers. In his book “The Conscious Mind” (Chalmers, 1996), David Chalmers calls the problem of subjective experience the hard problem, contrary to the easy problems which can be solved by standard methods of cognitive science. How can we bridge the gap between objective, physical reality and subjective experience?

John Searle gives in his book "The mystery of consciousness" the following formulation of the problem: "How is it possible for physical, objective, quantitatively describable neuron firings to cause qualitative, private, subjective experiences?" (Searle, 1997).

Can we explain the redness of red or the blueness of blue? How does a fur feel if we touch it, how does a freshly baked bread taste if we eat it, how do old books smell? Well, if someone wants to know what it is like to touch a fur or to eat a fresh bread, he can try it. To show someone what redness feels like we can therefore simply show the person the red object.

A physical explanation of color would be that red light has a wavelength around 620 to 750 nm, green light around 490 to 570 nm and blue light around 450 to 500 nm. Red light has the longest wavelength and blue the shortest. The physical explanation is easy. The subjective feeling of color in an individual is more difficult.

- We experience the same kind of pain as other humans because we have genetically the same A-delta fibers and C-fibers.
- We experience the same kind of pleasure as other humans because we have genetically the Endorphins and Dopamine molecules.
- We experience the same red, green and blue colors at different wavelengths as other humans because we have genetically the same cone cells for color reception (unless we have red-green color blindness).

We do not experience the world in the same way because each of has taken a different path and lives in a different location. People in China for example perceive the colors red and white different people in Europe. In China red symbolizes happiness, good fortune, vitality, and prosperity. Red is associated with weddings, while the color white is linked to mourning, death and funerals.

In Europe the color white is associated with weddings and peace flags, while red is the color of love and passion. Red is used as a color in warning signs to signal danger. The perception of these colors differs in both cultures.

Like every culture, every individual perceives the world a little bit different because everybody has experienced different painful and pleasant situations along the way. We tend to avoid what was painful and repeat what is pleasant. This leads naturally to path-dependence. Subjective experience is path-dependent.

A concrete physical mechanism of path-dependence works is named "preferential attachment". It describes the "rich get richer" network process where where well-connected nodes in a growing network are more likely to attract new connections and was popularized in network science by Albert-László Barabási and Réka Albert (Barabási, 2016).

The subjective feeling is based on the physical structures, but on top of the raw experience there is a learned aspect. Every organism from the same species that has the same photoreceptor cells in the retina to detect these colors in bright light will have a similar raw experience - unless they have a genetic defect which leads to different receptors as for example in red-green color blindness. As long as the person has the same photoreceptor cells the raw experience of redness, greenness or blueness will be the same. On top of this raw experience, we have personal memories and aspects we have learned over time, for example for blue the memory of blueberry ice-cream or a blue swimming pool on a hot summer day.

The answer why subjective experience is different to each of us is that we all have experienced a slightly different slice of the world, which led to a slightly different wiring and relationship structure in our society of mind. Each of us has a slightly different network of friends. Similarly, the agents in the society of mind for each person have slightly different friends, which depend on the encounters a person has made during the course of life. These friendship relations are different for each person. A thought is always associated with other thoughts, which are often related to things we have experienced at the same time or place in the past. If we want to understand the behavior and personality of a person, we must look at the curriculum vitae and the individual history of the person, i.e. we must consider the specific contexts and the various environments which lead to specific brain structures by continuous adaptation.

The question how subjective experiences arise from objective brain processes alone is wrong, because brain processes are not objective. Brain processes and neural connections are always unique, personal and individual, they are slightly different for each person. There are no two persons who have the same neural connections or brain processes. Everyone has a unique "society of mind". Someone who lives in Europe will have a brain that's wired up differently from somebody who lives in Asia or America. Each of us has different roots, has learnt different languages, has seen different places, has travelled along different paths, and has done different things. Because already Aristotle knew that we are what we repeatedly do, we are all a bit different, because we have all done things which are a little bit different and result in different memories. If all members of a species would be identical in genotype and phenotype, they would experience things in the same way. If our reactions would be hard-wired, then our behavior and our reactions would be similar, and we would feel in the same way, depending only on the physical constraints (for example that colors and tones correspond to different wavelengths, an "A" has 440 Hz, the color blue has 480 nm, etc.). But we are not hard-wired. Someone in China may associate the color red with good fortune, luck or joy (or simply with his country and his home). Someone in America may associate the color red with Heinz Ketchup and evil communists. Our behavior is based on a long process of socialization, assimilation, and adaptation. Adaptation is the answer how physical items can turn into subjective objects. Each of us is adapted to a slightly different world (or different "slice" of the same world, as Nicholas Thompson used to call it).

This agrees with what Sociologists and Psychologists say. For the sociologist Erving Goffman (1922-1982) the processes that cause subjective experience were clear, they were social, and they were associated with the own identity. Ego-identity is for Goffman "the subjective sense of his own situation and his own continuity and character that an individual comes to obtain as a result of his various social experience" (Goffman, 1963).

One recent book on the subject, "Soul Dust: The Magic of Consciousness" (Humphrey, 2012) from Nicholas Humphrey follows a similar line of thought. Phenomenal consciousness and personal, subjective experience is not something which only humans have. Every animal has a special subjective experience. Especially rich is the experience of course for intelligent and playful mammals like dolphins and the great apes. The subjective present of humans is particularly rich and deep, as Nicholas Humphrey argues in his book "Soul dust". It is full of colors and flavors, but each people sees different colors and feels different flavors. For a cognitive animal, every mental representation of the current situation has a personal dimension, it is colored with particular emotions and linked to personal memories.

The “colors” represent if the degree to which the events are perceived as pleasant and good or unpleasant and bad.

Some philosophers like Colin McGinn argue that the mind-body problem and the hard problem of consciousness are permanently unsolvable by humans (McGinn, 1990). We do not want to give up so soon.

Origins of subjectivity

This solution to the hard question has already been found or anticipated by William James and his scholar Edwin Holt. William James argues his text “Does ’Consciousness’ Exist?” that subjectivity arises from the different contexts in which we experience things. The pure subjective experience of items in the present depends on the mixed experiences with these items in the past, which in turn depend on the various contexts in which the objects and items appeared. These different contexts shape the relations between the individual experiences for each person. He says about the problem:

> “subjectivity and objectivity are [..] realized only when the experience is ‘taken,’ i. e., talked-of twice, considered along with its two differing contexts respectively, by a new retrospective experience, of which that whole past complication forms the fresh content. The instant field of the present is at all times what I call the ‘pure’ experience.[…] The peculiarity of our experiences, that they not only are, but are known, which their ’conscious’ quality is invoked to explain, is better explained by their relations – these relations themselves being experiences – to one another.”
> ~ William James, “Does Consciousness Exist?”

His scholar Edwin Holt goes one step further and shows us where these particular relations come from. For Edwin B. Holt, consciousness or mind is “a cross-section of the universe, selected by the nervous system”. Edwin B. Holt (1873-1946) was a professor of philosophy and psychology at Harvard and Princeton. Together with Ralph Perry he belonged to the New Realists, a group of William James’ students and associates. The New Realists were Edwin B. Holt (Harvard University), Walter T. Marvin (Rutgers College), William Pepperell Montague (Columbia University), Ralph Barton Perry (Harvard), Walter B. Pitkin (Columbia) and Edward Gleason

Spaulding (Princeton University). Together they wrote a book named "The new realism" (Holt, et al., 1925). In this book Holt says about the cross-section:

> "It must not be forgotten that while the object itself, if a physical thing, is far from simple, we are always perceiving it in a complicated setting of (spatial, temporal, and logical) relations, which is a still more complicated thing. But the conscious cross-section is always a group of the integral (neutral) components of the object and of its innumerable relations. [..] it is seldom possible to say just where the object itself terminates and its relations to other entities commences"
> ~Edwin B. Holt

He defined a cross-section in general as a definable part of a larger collection or manifold. Each of us experiences a specific cross-section during the course of life. In his book "Concept of Consciousness" (Holt, 1914), Edwin B. Holt says about this cross-section (on page 171)

> "The sum total of all the whales living in certain given waters is a cross-section of the sea that is significant for the whalers who are trying to locate and gather them in. The various shafts and levels of a mine are a cross-section of the mountain, and of import to the shareholders: and it is the business of the engineer so to direkt the workings that this cross-section shall coincide with taht other cross-section that is made by the vein of ore."
>
> "Once again, a navigator exploring his course at night with the help of a searchlight, illuminates a considerable expanse of wave and cloud, occasionally the bow and forward mast of his ship, and the hither side of other ships and of buoys, lighthouses, and other objects that lie above the horizon. Now the sum total of all surfaces thus illuminated in the course, say, of an entire night, is a cross-section of the region in question that has rather interesting characteristics"
> ~Edwin B. Holt

Similarly, the mind selects all "elements or parts of the universe" which "matter", which are interesting or arise emotions, which cause the system to make a specific response. The mind is made up of all experiences the person has made during the course or path of his life. The experiences which form the mind are path-dependent.

Positive experiences with a subject reinforce other positive experiences in the future. If we like someone or something then we may get rewards in form of positive feelings from it, and the more good feelings we have in turn about something, the more we like it. A positive feedback loop arises, which amplifies slight differences into large variations. Each mind corresponds to a specific, individual slice of the same world. It is this cross-section which defines our subjective experience and makes our experiences individual.

If we really want to understand subjective experience, we must examine the past of a person. Maybe one could call it "mind archaeology" – digging through the layers which record the past of the individual. Of course this is not new at all, it would be very similar to psychoanalysis. Already Freud's patients lay on the couch during psychoanalysis and told him their stories and dreams.

The idea that the mind is composed of layers which recorded the past of the individual goes back to the 19th century. Julian Jaynes observed in his book "The Origin of Consciousness in the Breakdown of the Bicameral Mind" (Jaynes, 1976) that the metaphors used to understand consciousness changed over time, because they followed current trends in science. Today in the age of web 2.0 and social media, we tend to use social metaphors like the "society of mind" or the "social network of the mind". In the 19th century, physical sciences like geology were very popular. Jaynes says in "The Origin of Consciousness":

> "The first half of the nineteenth century was the age of the great geological discoveries in which the record of the past was written in layers of the earth's crust. And this led to the popularization of the idea of consciousness as the being in layers which recorded the past of the individual, there being deeper and deeper layers until the record could no longer be read"

John R. Searle writes in his book "The Rediscovery of the Mind" (Searle, 1992) that subjectivity is the feature of consciousness that is the most puzzling to philosophical analysis. He writes on page 95 "we find it difficult if not impossible to accept the idea that the real word, the world described by physics and chemistry and biology, contains an ineliminably subjective element. How could such a thing be?". An interesting question. The answer is probably just path-dependence.

We have argued earlier that each of us has taken a slightly different route during his life and is adapted to a slightly different world, i.e. everyone sees a different picture of the world on his trail and has experienced a different “slice” of the same world along his path. Our subjective experiences arise from a long path-dependent learning process. The pure experience of the present depends on the whole series of previous experiences in the past, our subjective judgements depend on the whole series of previous judgements and personal evaluations (is it good or bad for me?).

Peter F. Strawson (1919-2006), English philosophy professor at the University of Oxford, writes in his book “The Bounds of Sense” (Strawson, 1966) that a series of experiences has a double aspect. It contributes to our shared understanding of the common objective world, and at the same time it extends our individual subjective experience of that world. He says on page 105 (chapter “Unity and Objectivity”):

> “On the one hand it (a series of experiences) cumulatively builds up a picture of the world in which objects and happenings (with their particular characteristics) are presented as possessing an objective order, an order which is logically independent of any particular experiential route through the world. On the other hand, it possesses its own order as a series of experiences of objects. If we thought of such a series of experiences as continuously articulated in a series of detailed judgements, then, taking their order and content together, those judgements would be such as to yield, on the one hand, a (partial) description of an objective world and on the other a chart of the course of a single subjective experience of that world.”

All members of the same culture learn the same public language through supervised learning, but the results of unsupervised learning controlled by personal emotions is private and individual. The former is not path dependent, because supervised learning depends on an external teacher and does not depend on the own judgement. The latter is indeed path dependent, because it involves the self and is affected by positive feedback: small initial disturbances are reinforced and can cause large effects. In this sense subjective experience is self-referential: current positive/negative experiences depend on previous positive/negative experiences, etc.

This means adaptation occurs in two flavors: a shared, objective during common supervised learning (for example in schools, colleges or universities), and a personal, subjective during unsupervised learning. The former is responsible for our common understanding of the public world and the latter is responsible for our subjective experience. Strawson continues to say that this double aspect of adaptation and experience can be found even in the smallest unit, the raw experience itself, and becomes visible if a judgement of an experience is altered by hindsight:

> "Not only the series as a whole, but each member of the series, has a double aspect. This explicitly emerges when one objective judgement is corrected by another: what remains unaltered when the correction is made is the subjective experience."

The history of a normal human being determines both, the objective and the subjective experience. Path dependent subjective experience arises from the individual, unique, and personal route one takes, one among many other possible routes. The objective experience, i.e. the common understanding of our shared knowledge, depends more on the goal and the milestones of the route, than on the actual path taken. We see the same world, but along different routes, and from different angles.

As Searle has noticed (Searle, 1992), every subjective experience is always someone's subjective experience. It is associated with a certain path taken by a particular person. It always belongs to someone, and can only be validated by that person, whereas objective facts and knowledge belong to everyone, and can be validated or checked by everyone.

	Subjective experience	**Objective knowledge**
Relation	Involves the self	Does not involve the self
Ownership	belongs to someone	belongs to everyone
Acquisition	by unsupervised learning	by supervised learning
Measurement	needs a jury (emotions)	can be measured
Dimension	good-bad (psychological)	big-small, fast-slow (physical)

Understanding an individual route which leads to a specific subject experience is of course difficult, but it is much less complicated if we also consider the environment. Herbert Simon has noticed in his book "The Sciences of the Artificial" that "an ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself." (Simon, 1970).

The purpose of subjectivity

As we have seen in the last chapter, subjective experience is path-dependent. W. Brian Arthur was one of the first who discussed the concept of path-dependence in complex adaptive systems. He described how minor historical accidents can "lock in" an inferior technology or standard (like QWERTY or VHS), making economic outcomes unpredictable and path-dependent. The result resembles a "frozen accident". He defines path dependence as a dynamic process where random, minor historical events are amplified by self-reinforcing positive feedback loops, locking the system into a permanent structure that cannot be predicted from its initial conditions. A small historical accident gives one option a slight edge, which snowballs until it becomes a permanent standard. In his book "Increasing Returns and Path Dependence in the Economy" he describes that the same concept can also be applied to human learning. A mind can quickly "lock in" to cognitive rules, habits, or beliefs in the same way that the economy locks in to technologies and techniques (Arthur, 1994).

Path dependence is a general principle that can also be applied to different complex systems like political and social systems. Paul Pierson for example applied the idea from Arthur to political systems to explain why the United States is stuck with the Electoral College and Europe with the EU in Brussels (Pierson, 1994).

Subjective experience does not only depend on the personal path of the individual, but also on the history of the large social groups he belongs to. Pierre Bourdieu was one of first who observed that cultural taste and aesthetic choices are deeply shaped by social factors and serve as a weapon in the struggle for social status. The personal taste that someone has normally fits to the class or tribe that he belongs to - or wants to belong to (Bourdieu, 1987).

Carl Wilson also argued 30 years later that our social, economic, and cultural background and our regional origin shape our taste. We are trained from young years to find certain things comfortable and others not. Our subjective experience is path-dependent (Wilson, 2014).

In his book “Status and Culture”, W. David Marx argues that our tastes and what we like are driven by a fundamental human desire to signal the social group we want to belong to and the identitiy we would like to have. We adopt specific behaviors to show which groups we belong to and how "high" or "low" we stand in society (Marx, 2022).

Wearing specific brands (sporty, luxurious, adventurous, smart, etc.) helps us to archieve a certain social standing. We feel good wearing a specific brand because it helps to confirm our belonging to a certain social group.

- If we want to be rich we probably wear clothes of a luxurious brand that signals we are rich
- If we want to be smart we might wear clothes which have a university logo that signals we are smart
- If we want to be sporty we will wear clothes of a sporty brand that signals we are sporty

The brand validates what we want to be. In this sense the brand we decide to buy indicates what we are not. We could say our subjective experiences and preferences are the visible surface of our invisible life story. We tell others who we want to belong to and what we want to be seen as, just as Erving Goffman observed (Goffman, 1956).

A person's personal taste usually corresponds to the class or group to which they belong—or to which they wish to belong. Just as every person has an own path dependent history, every culture or tribe has its own path dependent history. We get a self-similar path dependency. On the macro-scale the nation or culture, on the meso-scale the tribe or family and on the micro-scale the individual person itself.

Given that there is so much path dependency it might be surprising that we can agree on a shared understanding of the world at all. Objectivity is the residue that remains invariant if we switch perspectives - perspectives that are distorted individual preferences, idiosyncratic descriptions and personal history.

Nozick argues in his book “Invariances” (Nozick, 2001) that subjectivity depends on our point of view, while objectivity does not. The subjective view is linked to subject, while the objective view is not. Thomas Nagel called the objective view "the view from nowhere" (Nagel, 1986).

Subjectivity would be a highly functional, evolutionary adaptation that arises naturally within the objective world (Nozick, 2001). The evolutionary advantage is fast decision making that is aligned with the primary directives of the organism (survival and reproduction). In a sense our experience "has to be subjective", i.e. colored by subjective judgement if it is good or bad for us, because only then we can make decisions which are fast enough in eat-or-be-eaten world of prey and predators.

The best way to understand the difference between subjectivity and objectivity is in terms of perspective. The default perspective is the subjective experience. Subjectivity is the way in which a subject organizes the own daily "stream of reality". An organism naturally carves up reality into objects and things, and evaluates them if they matter or not, if they are good or bad for it. Subjectivity emerges naturally from this 1st-person view.

- Subjectivity is a single unique perspective shaped by personal views, feelings, beliefs, or opinions (the special view from somewhere)
- Objectivity is a view of the world which is invariant by change of perspective. This is essentially the view from nowhere, as Thomas Nagel called it (Nagel, 1986), or the view from everywhere

To obtain an objective view we need to detach ourselves from a specific perspective or 1st-person view, as Thomas Nagel describes in his book "View from Nowhere" (Nagel, 1986). The things that remain persistent patterns over time are called objects, the objects that remain invariant if we change the perspective become part of our objective view.

Subjective experience
↓ invariant to change in time
persistent pattern / object
↓ invariant to change in perspective
subjective, evaluated object
↓ invariant to change of observer and location
intersubjective, neutral object

↓ invariant to change of instrument
Objective experience by using objective model

Objectivity is the objective view that remains invariant under all transformations of perspective: when the viewpoint is transformed or the perspective is changed or the observer is replaced. Since emotional valuations are associated with a certain perspective, this objective experience must be free of emotional valuations.

Paul Thagard argues that all conscious qualia are fundamentally shaped by emotional valuation (Thagard, 2025). He proposes the following "equation of qualia":

Qualia = Neural Representation + Cognitive Appraisal & Physiological Perception

By combining the what component (what does it represent) with the how components (how important is it for me & how does my body react) mind and body are getting ready for action.

The reason is simple: our feelings help us to recognize quickly what is important or what is not. Subjective experience and emotions evolved because organisms needed rapidly action-guiding evaluations of the world rather than neutral representations. Evaluating things by feelings as good, bad, dangerous, attractive, painful, or rewarding compresses biologically relevant information into immediately usable form. In an evolutionary environment of predators and prey, such perspectival valuation enables adaptive behavior quickly enough for survival. The drawback is that the information we perceive is always a little bit biased.

The subjective aspect of conscious experience is tied to the evolutionary need for fast, adaptive decision-making. A gazelle does not neutrally register: "Big cat approaching at high velocity" or "A massive object composed of muscle and claw is moving toward my coordinates at 40 km/h" and then starts to think what it could do. It experiences an urgent sense of imminent danger, panic and fear which makes it ready to run for its life immediately.

Without subjective experience and without qualia nothing would have value or valence, and we would not be able to be able to react quickly to novel, unpredictable situations. Nozick uses the term Flexible Response Mechanism (FRM) for a system that can respond to changing situations in multiple context-sensitive ways rather than through fixed stimulus-response patterns (Nozick, 2001).

Inner and outer world

While there are good reasons for subjectivity it leads to many isolated first-person perspectives. In his book "Putting Ourselves Back in the Equation: Why Physicists Are Studying Human Consciousness and AI to Unravel the Mysteries of the Universe" (Musser, 2023), George Musser defines the inside/outside problem as a fundamental tension between the third-person description of the objective world that physical science traditionally seeks to explain and the first-person perspective of an embedded, subjective observer which is only available to the observer himself.

How is the inside/outside problem related to the classic distinction between physical reality (matter) and mental ideas (mind)? This is a deep question philosophers have discussed since Descartes ("I think, therefore I am") for centuries. Descartes' dualism says that the mind is a non-physical, thinking substance, and the body a physical, extended substance. They would be two fundamentally different things, yet able to interact.

Philosophers like Gilbert Ryle argued there is no such thing as two substances, no ghost in the machine. But we certainly have a distinction between the matter in our head, which we can not easily access or observe, and the matter in the outside world, which we can study and measure objectively. The points of interaction are of course simply the sense organs which transform the light and sound waves that describe the outside world into electric pulses and patterns inside our brains. And our muscles which convert electric pulses back into physical movement. The mind is obviously related to all the inside stuff, encoded in the neurons and synapses inside our brains, and the physical reality is related to the outside objects.

The hard problem is to cross the boundary between inside and outside, between the interior world of subjective experience and the exterior world of physical objects. It means to try to understand the first-person view in others (the hard problem of subjective consciousness which we have discussed a number of times) or the third-person view of ourselves (related to self-awareness and the problem of free will, which we also discussed in the last chapter). From this point of view the two fundamental problems of the hard problem of subjective consciousness and the problem of free will are related, because we have to cross the inside/outside boundary in both cases.

Explaining and understanding our inner experience requires crossing the inside/outside boundary of inside experiences and outside objects. Language in spoken and especially in written form that our minds are able to master after years of learning and school education allows us to cross the inside/outside boundary. No new physics required. Only new machines to make it easier to cross from one side to the other: we know them well. First we witnessed the emergence of writing systems in ancient civilizations, then the first books, the first novels and early theaters, after the invention of the printing press many more books, and eventually cinemas and TV screens. They all allow us to share our experiences and to synchronize our knowledge of the world.

Cinema and TV screens can be considered as the windows where we can observe the actions and perceptions of someone else. They use technologies specifically designed to make the first-person perspective available to a third-person view. They are also frame merging or frame sharing devices: for the time of the movie we and the leading actor experience the world through the same set of synchronized frames. This allows us to transcend the inside/outside boundary that separates the inner world of subjective experience and the outer world of physical events.

Cinemas make the worlds of others accessible to their viewers. They are fascinating for us because normally the character and personality of a person is not observable. It is only known by family and friends. In this sense a personality is a windowless substance, as Leibniz called it. Cinemas open windows which let us peer into the life of others: we start to understand their subjective experience and witness how their personalities and characters are formed by witnessing how they react to challenges and difficulties.

They have this magic capability because they are able to cross the inside/outside boundary mentioned in the beginning. Film cameras capture like sensory organs physical forms and their movements, transform them into electric pulses and patterns, and store them in encoded form in some kind of medium. Film projectors like motor organs and muscles transform the patterns back into physical forms and their movements.

Show business and film industry

The hard problem of consciousness is related to nothing else but the show business and film industry in Hollywood and elsewhere. Why can a cinema or a movie theater be seen as a machine to understand the profound and fundamental question of what it is like to be someone else?

To use the words of Roger Ebert everybody is born unwillingly inside the movie of his own life. Each of us is stuck inside his or her own movie. Roger writes in his memoir "I was born inside the movie of my life. The visuals were before me, the audio surrounded me, the plot unfolded inevitably but not necessarily. I don't remember how I got into the movie, but it continues to entertain me". In short we are the main actors and spectators in the movie of our own life.

Apparently, cinemas allow us to view the movie of other lives. We are fascinated by movie theaters and cinemas because everybody is stuck in his own movie, and they allow us to see how the other movies look – or in other words what it is like to be someone else. While the movie of someone's life is not the real life, it brings us as close as we can get to it, especially if it is well made. Roger Ebert said "Art is the closest we can come to understanding how a stranger really feels."

Vivian Sobchack argues that a film in a movie theater does not just display reality – it presents a world that invites our embodied participation. She says in her book "The Address of the Eye" (Sobchack, 1992) that for a brief moment "we forget ourselves in our interest in another's vision of the world". By perceiving the vision of the world from someone else they bridge the gap between our own subjective "movie" and the "movies" of others. She argues the film experience provides empirical insight into the nature of that embodied vision we each live daily as both "mine" and "another's.

> "The cinema thus transposes what would otherwise be the invisible, individual, and intrasubjective privacy of direct experience as it is embodied into the visible, public, and intersubjective sociality of a language of direct embodied experience."
> ~ Vivian Sobchack in "The Address of the Eye"

And this is what the hard problem of consciousness is all about: understanding how a stranger really feels. There might be no single Cartesian theater where consciousness happens as Daniel Dennett has argued, BUT theaters are the place where the hard problem of consciousness is solved. A movie theater or cinema is a machine built to solve it. Not perfectly, but approximately. Like a numeric solution of an algebraic or differential equation it gives not an exact solution, but an approximate result which helps us to understand how the solution looks like.

Mathematically we could write it as the limit of the qualia function f_q

$$\lim_{d \to 0} f_q(d) = Q$$

..where Q is Qualia and d is distance between 1st person views - the topological or structural difference between two path-dependent graphs. It represents the difference in path-dependent architecture and emotional valuation matrices. As d approaches 0 the qualia of one person converges towards the qualia of the other person (e.g, two humans with identical cultural and family backgrounds or two viewers of the same movie)

A cinema is able to do something similar because it…

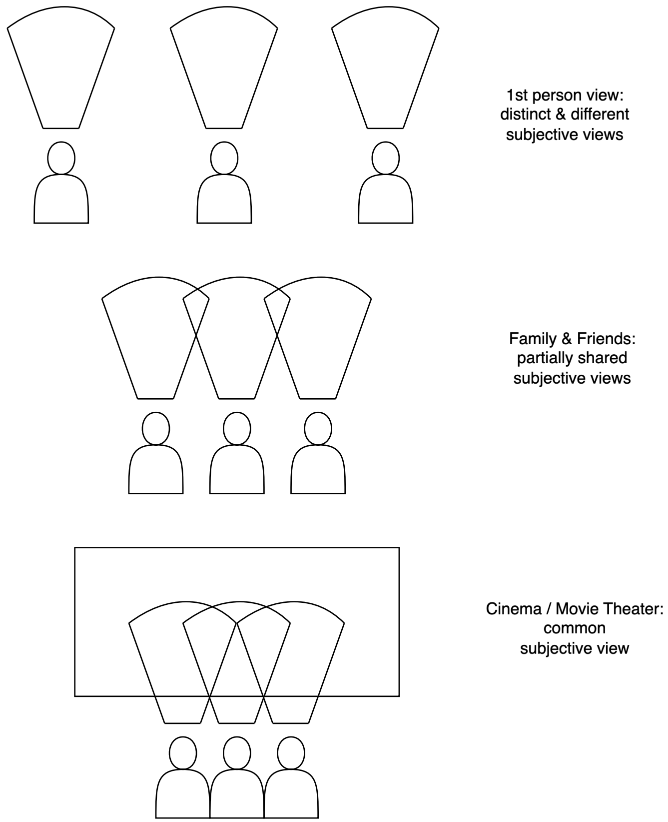
- **changes our perspective**:
  We see the world from the point-of-view of somebody else, through the eyes of the actors, and this allows us to share their emotions. We feel their fear, joy, sorrow, and confusion. Cinemas bring us as close as possible to the subjective reality of other persons. They allow us to bridge the gap of subjectivity by giving us the opportunity to perceive the experiences of a 3rd person from the 1st person point of view. Thereby we can experience a subjective perspective that the real world normally does not offer.
- provides us **a window into different worlds**:
  Cinema allows us to step into the shoes of people from different cultures, socio-economic backgrounds, historical periods, or even future times. It provides a window into “what it’s like” to be a refugee, a hero, a soldier, a researcher or a scientist, someone struggling with an illness, a person of a different race or gender, or someone facing an impossible moral dilemma. These are experiences we might never have in our own live
- enables **immersive storytelling**:
  The physical setup of a cinema – the darkened room, the large screen filling the field of vision, the powerful sound which surrounds the viewer – is designed to minimize external distractions and maximize immersion

> in a different world, the world of the main actor in the story that is being told. By understanding the story step by step, we understand what it is like to be someone else. A cinema allow us to step into someone else's shoes, and to follow his or her path

Can we solve the hard problem of consciousness, i.e. understand the subjective experience of others? Subjective experience is all about the "What it's like" aspect which is not accessible to others. Although we sometimes feel the pain of others we have in general no access to their inner experience from the outside. We cannot measure or observe it directly. Only indirectly by using the machines we have built to do so.

Cinemas are machines we have built to experience what it is like to be someone else: a hero, a winner, an adventurer, a musician, etc. Since subjectivity is path-dependent we can try to narrow the gap between different first-person perspectives by following the path of the person in question. Cinemas and movie theaters are tools we have built to do this. They give us for a short moment the feeling what it is like to be someone else. It is not the same as the real subjective experience, but it is approximately the same, depending on how convincingly the story is told and how deeply we follow it.

Subjective experience is something we cannot share directly. Only indirectly. Since it is path-dependent we can basically follow the same path to experience similar situations. Robert Nozick would say we can "track" it. Understanding subjective experience can be an art. It is a form of consciousness, but not of the self. It is a consciousness of others where we transcend the boundary of 1st person view and 3rd person view.



If we try to summarize the solution to the hard problem we have sketched so far then it does not sound that hard:

Why should any physical process be accompanied by subjective experience at all? Why is there a first-person subjectivity at all? Why is there something it is like to be a system at all? Because emotions and conscious experience in general provide an evolutionary advantage, as Nozick says. Subjective experience and emotions evolved because organisms needed rapidly action-guiding evaluations of the world rather than neutral representations (Nozick, 2001).

How does objective physical structure become subjective experience? How does subjective experience arise? Emotions help to evaluate each new situation instantly. They attach a value or valence to every event. Depending on the outcome of each situation, the valuation of events and objects are subject to path-dependent reinforcement learning.

How can we bridge the gap between two different first-person perspectives? By using the machines built to bridge it. The machines from show business and movie industry which let us track the path of the protagonists can close the gap, but not completely. Only temporarily and approximately. Understanding subjective experience is not impossible, but it is an art. To be more precise: a whole art form genre.

What remains hard is to bridge the gap between different subjective experiences. We can only close it approximately, not completely. Which is the reason we have built complex machines that make lots of money solving the problem. And the reason why it is named "the hard problem".

Philosophy professors would probably find difficult to admit that film industry and show business make lots of money every day by solving the hardest philosophical problem they have tried to solve in vain for centuries. But apparently this is the way it is.

References

+ John Searle, Mind - A Brief Introduction", John Searle, Oxford University Press, 2004
+ Gilbert Ryle, The concept of Mind, University of Chicago Press, 1949
+ Roger Ebert, Life Itself: A Memoir, Grand Central Publishing, 2012
+ David Chalmers, The Conscious Mind: In Search of a Fundamental Theory, Oxford University Press, 1996
+ Daniel Dennett, Consciousness Explained, Little, Brown and Company, 1991
+ Robert Nozick, Invariances - The Structure of the Objective World, Harvard University Press, 2023
+ Vivian Sobchack, The Address of the Eye: A Phenomenology of Film Experience, Princeton University Press, 1992
+ George Musser, Putting Ourselves Back in the Equation: Why Physicists Are Studying Human Consciousness and AI to Unravel the Mysteries of the Universe, Macmillan Publishers, 2023

The mind-body problem - the paradox of the ghost in the machine

Executive summary of core points:

- The ghost in the machine paradox: we seem to have a mind that is different from the body, and yet we don't have it because we are made of physical matter. Is there a ghost in the machine or not? How can a physical system contain a separate, non-physical entity as a mind, a soul or a ghost?
- Elements interact with other elements in their system and one level above or below, patterns interact primarily with other patterns on the same level of abstraction. We cannot explain high-level patterns by low level elements if the distance is too large
- As Aristotle said the body is made of matter and the mind the pattern that organizes it

The philosophical theories

> "Minds are not merely ghosts harnessed to machines, they are themselves just spectral machines"
> ~ Gilbert Ryle (1900-1976), in "The Concept of Mind" (Ryle, 1949)

If we ask a philosopher whether the famous philosophical mind-body problem exists the answer would "Yes, of course. It is one of the biggest problems". The Oxford companion of philosophy writes "one thing that is certain is that the mind-body problem is one of the deepest puzzles in philosophy, and that it will continue to test our philosophical intelligence and imagination" (Honderich, 1995).

If we ask a neuroscientist like Candace B. Pert, the neuroscientist who discovered the opioid receptor for endorphins (Pert, 1997), whether the famous philosophical mind-body problem" exists the answer would be "No, of course not. The mind is simply what the brain does" or "No, certainly not. Just as traffic emerges from a collection of moving cars, mind emerges from the collective activity of billions of neurons".

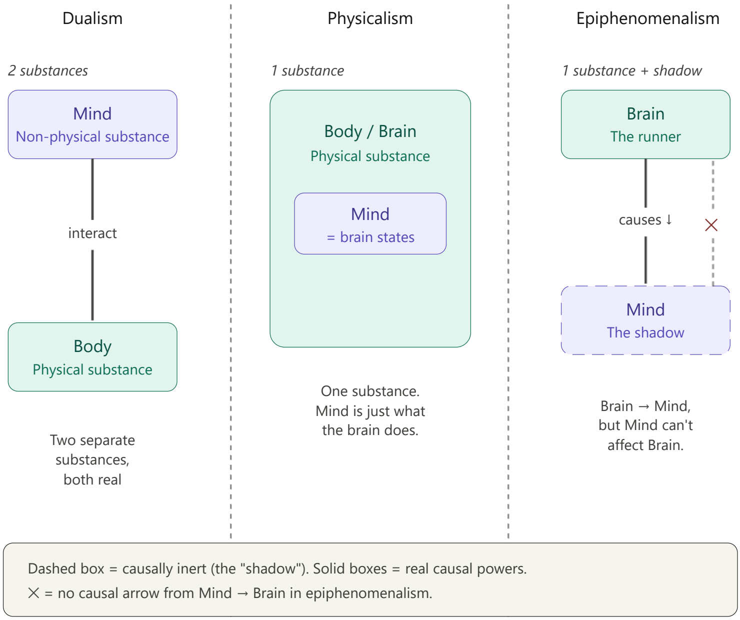
The philosopher who lives in the world of language says there is a mind-body problem, and the neuroscientist who lives in the world of molecular biology says there is no such problem. Both fail to see the world of the other. The philosopher ignores molecular biology, biology and biochemistry while the neuroscientist ignores the mental world of the mind: the world of language.

Apparently, the relationship between mind and matter is complicated. The ancient Romans used to say "Mens sana in corpore sano" ("A healthy mind in a healthy body"). German poet Friedrich Schiller wrote in his play Wallenstein “It is the mind that builds the body” (“Es ist der Geist, der sich den Körper baut”) and emphasizes the power of mind over matter. According to Schiller the mind in the form of attitude and mindset is able to shape the body.

Marvin Minsky wrote in his "Society of Mind" book that "the mind is what the brain does" (Minsky, 1986). Some claim the mind builds the body, others that the body builds the mind. Who is right?

The three big philosophical theories to describe the mind-body problem are called

- Dualism (two substances, mind and body)
- Physicalism (one substance, everything is made of physical substance)
- Epiphenomenalism (one substance and a "shadow". The mind is like the shadow cast by a runner. The runner/brain causes the shadow/mind, but the shadow doesn't help the runner win the race)



Are mind and body distinct substances? This has been debated since ancient times. Descartes argued yes, others like Ryle were convinced the answer was no.

The ghost in the machine

The phrase "The Ghost in the Machine" was coined by Gilbert Ryle to mock Descartes' Dualism. Ghost buster Ryle argues there is no ghost in the machine (Ryle, 1949). He says the mind is just the name for the ways a person behaves and responds based on their capacities, the collective activity of the whole person interacting with the world. The mind would not be a hidden "ghost" inside the physical structures of the body. He uses a university analogy to illustrate this. If you walk through Oxford, you see several different libraries, lecture halls, labs, colleges and student houses, but no additional hidden building. Ryle argues asking “where is the university in Oxford” would be wrong because there is no additional, hidden building for the university itself, only the collection of everything you see if you walk through to the city. Similarly, it would be a mistake to suspect a hidden ghost inside the brain.

I believe Ryle threw out the baby with the bathwater by dismissing the “ghost” as a mere linguistic error or a category mistake. The interesting thing about the paradox of “the ghost in the machine” is that the ghost in the machine is real, and at the same time it is not.

We have no evidence that immaterial ghosts harnessed to biological machines or spirits coupled to biological systems exist. Unless we think of immaterial as abstract in the sense of Platonic ideas. The form of the overarching organization, i.e. the personality and the character of a person in combination with all the levels below that implement it, can make the impression as if there is a ghost in the machine.

It looks as if there are ghosts harnessed to machines, but in fact the ghost emerges from the machine itself. In this sense the “ghost in the machine” emerges together with the personality or the character of a person, and describes the overall abstract organizational pattern of the individual behavior, similar to overall organization of the University of Oxford that Ryle used as an example.

If a ghost is simply the mental representation of a non-existent object, our construct of an "inner self" fits the definition of a ghost if such a self does not exist. In moments of self-awareness, we actively create this imaginary self. While it does not physically control us, this "ghost" lives with us and comes alive during moments of self-awareness.

In the chapter about self-awareness, we have seen that the ghost in the machine is actively created during moments of self-awareness. The argument went like this: Philosophers like Gilbert Ryle and Daniel Dennett argued against the idea of a unified, inner "self" controlling the body from a single command center in the brain (Dennett, 1991). But Dennett actually proves the ghost in the machine **does** exist precisely because there is no real self.

Quantum Gravity of the Soul

If we look closely then on the lowest level every little thing is indeed clearly physical because everything is made of physical parts, of atoms and fundamental particles. On higher levels of abstraction things look different. The atoms in chemistry and molecular biology still follow the laws of physics but they follow their own rules if they are aggregated in molecules.

Like everything else we are composed of blind and deaf matter that follows on the lowest level strict laws of physics, yet we experience sight and sound - a subjective, internal world of color, pain and pleasure, dreams and hopes. How can a purely physical system (the brain) produce a non-physical experience (consciousness)? We are a part of the universe that is at the same time somehow "apart" from it, observing the very laws that govern us.

One possible answer is to deny that mental phenomena and the mind are non-physical. John Heil (2015) says in "The Universe As We Find It" that "The mental and the physical are names, not of families of substances and properties, but of ways we have of conceiving, describing, and explaining the universe." He argues that the distinction between “mental” and “physical” is not fundamental because both are just different ways of describing the same underlying properties, two different perspectives on one underlying system. There would only be one kind of "stuff" in the universe — physical substance described by fundamental physics. Heil says what we call "mental states" are just physical properties of a complex system (the person) described at a certain level of abstraction. He argues there is only one level: the fundamental level. All "higher-level" sciences are not describing new "emergent" things; they are simply describing the fundamental things using different levels of precision or abstraction.

Let us put it like this: as we climb up the ladder of abstraction (elementary particles, atoms, molecules, amino acids, DNA, neurons, cells, organs, organisms, …) new patterns and structures emerge which interact in new ways that can be described by new symmetries, i.e. they are governed by new laws which depend on the scale of observation. Each level and each system and their own elements which interact by their own laws and form their own substance. As Alwyn C. Scott has observed when systems become complex, the interactions between their components create entirely new laws and phenomena that cannot be predicted by looking at the pieces alone (Scott, 2007).

This means asking “How does a non-physical thought cause a physical action or movement if the mind is non-physical?” or “How does the mind emerge from mindless matter?” is basically asking to explain high-level patterns by low-level elements, and explaining highest level patterns by lowest level elements is just as futile as finding a theory of Quantum Gravity.

Feature	Quantum Gravity	Consciousness
The small scale	Quantum Mechanics (Particles/Fields)	Biology & Chemistry (Synapses/Neurons)
The large scale	General Relativity (Spacetime/Gravity)	Psychology & Sociology (Identity/Subjective Experience)
The missing link	We can not find a “Graviton” or unify the math	We can't find the "soul" or unify matter with mind
The paradox	Space is both "smooth" and "pixelated”	We are both matter and mind

Language, i.e. characters, numbers, words and sentences, math, symbols, and code are immaterial and abstract. They are abstract entities that need a physical medium or material to be stored, transmitted and read. Any deformable material can be used to encode them. Their meaning is not contained in the medium, rather in relationships to other abstract entities and patterns. A "5" is still a "5" whether it is carved in stone, written on paper or blinking on a screen.

Could the ghost in the machine paradox be related to the embedding and encoding of abstract patterns and rules? Aristotle argued in classical philosophy in his treatise “On the Soul” that the soul or psyche (ψυχή) is the form, or essence of any living thing. He defines a soul as that which makes a living thing alive. and says that a soul is related to its body as form to matter. Aristotle called this doctrine where everything is a combination of matter and form hylomorphism. If the "machine" is the matter, the "ghost" can be seen as the abstract set of patterns or forms that organizes that matter.

Could it be that the personality and character which we develop over time is an approximation of an abstract behavior pattern for a certain type of person? This type of person would be a complex high-dimensional pattern, form or shape in behavioral space.

At birth, that shape is a blurry sketch based on temperament and genetics. We start as young physical organisms with nothing but raw emotions. During the course of life we first learn to speak, to read and to write, and then we develop a certain personality which depends on our upbringing and education, our individual experience and preferences, our schooling and training.

As you live, every experience, decision, choice, and reaction adds more dots and edges which increase the resolution of your personality until it is fine-grained enough to be recognizable. As Aristotle said, “it is not one swallow or a fine day that makes a spring”. We are what we repeatedly do. This construct is like a digital image which is approximated by dots or a mosaic of small colored stones and becomes more fine-grained over time. Because it is expressed in language this type of person could also be realized by something completely different, like an AI agent, but it can also be encoded in neurons and synapses, i.e. ultimately in biological molecules.

Through education and learning, work and daily life we begin to realize a pattern that comes close to an abstract personality, something transcendent. By living in a way that obeys higher-level rules, for example rules related to boldness, kindness, friendliness or whatever personality we are trying to become, we can obtain a certain kind of character and personality. If it is recorded and described somewhere it can outlast and transcend the person. We could call this kind of dualism in a single system in the words of Jaegwon Kim "dualism, or something near enough".

Or *dualism by asymptotic emergence and entanglement*. Entanglement means here that two different systems are merged into one. All substances are made of physical substance at the smallest scale, and Jaegwon Kim’s principle of supervenience says there is no mental change without a corresponding physical change. And yet many elements of the mental system like words, codes, symbols and languages in general contain abstract relations that can also be engraved in stone or contained in books. The physical system is realizing abstract elements, patterns and processes from a different system, the "mental" system of language, codes, and symbols. Jaegwon Kim calls this principle "physical realization".

If you zoom in far enough on a human "personality," you don't find a "person" - zooming in the brain you find individual firing neurons, neurotransmitters, chemical signals, atoms, icons and electrons. You cannot see the forest for the trees. The "character" and "personality" only emerges when you step back and see the full resolution of the pattern. A high-level pattern like a general rule is not able to influence individual neurons, atoms or data points. More abstract patterns, like the rule "Do not kill others", can interact with and govern more concrete rules like "Do not eat animals, because this would be murder".

Therefore, I would argue we have a ghost in the machine, but the ghost is not a mysterious substance or matter, it is a structure or a pattern on a higher level of abstraction. It is the "personality" that emerges over time and which is embedded in high-level physical patterns that realize abstract rules. It can converge to an ideal type, but since we rarely manage to become the person we long to be perfectly, we can speak of asymptotic emergence.

We have complex cognitive material patterns that approximate an immaterial soul. These patterns include the complex patterns of consciousness during self-awareness which resembles a strange attractor that was discussed earlier and the complex rules of our own character and personality that we acquire during a lifetime of learning.

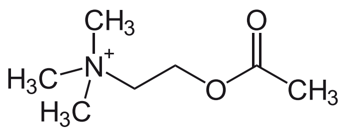
Bridging the gap

There are two different worlds we live in. Each world has its own fundamental elements and substances. We all learn to bridge the gap between them when we are young. There is a mental/symbolic world which consists of letters, words, meanings, thoughts and intentions, and a physical world which has sounds, movements, and physical bodies and physical activity.

On a physical level a thought is just electric current which zips along the billions of nerve cells as neuroscientist Hannah Critchlow would say (Critchlow, 2019). A brain contains 80-100 billion neurons each of them having on average 10,000 synapses which results in trillions of connections. In the neurons themselves the information is passed as an electrical signal along the axon to the synapses which connect neurons to other neurons. Between neurons neurotransmitters transmit the information from one neuron to another. It is a lightning-fast web of communication where electric current zips along billions of nerve cells.

In this picture the mind is a complex, shifting pattern of electrical energy. The electric current in the neurons is just following the laws of physics, chemistry and molecular biology and is guided by the plumbing of your brain, i.e. it depends on which neuron is connected to what other neuron by how many synapses. The wiring in turn depends on our path-dependent experience and on your character and personality which has developed over time.

The body is the billions of other cells we can see and touch: our skin, our bones and all the rest. One could argue the neurons themselves are both, mind and body, because they are the place where the mind is active and they are also part of the body. In this sense the neurons are the connection between mind and body. In combination with the sensory end points, the sense organs which transform physical events intro electric current, and the motor end points, the muscles which transform electric current back into physical events, they are the place where mind and body meet.



Neurotransmitter Acetylcholine $C_7H_{16}NO_2$ (from Wikipedia)

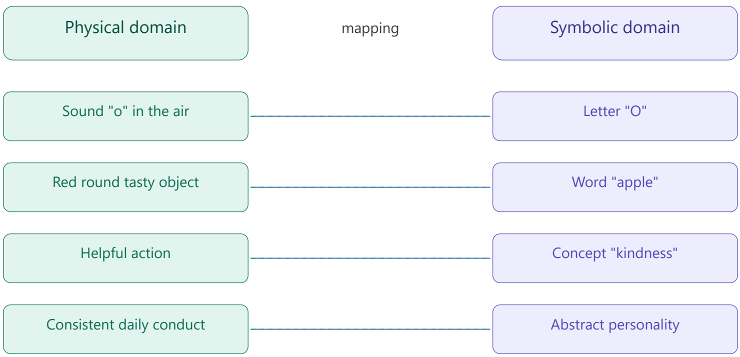
Neurotransmitters like Acetylcholine are used by neurons to control muscles. They can be found in the synapses of motor neurons to transmit information between motor neurons and muscle fibers. Since they connect the mind in form of neurons and the body in form of muscles, they are in a sense the ultimate connection of mind and body.

Neuropeptides are also used as a connection between mind and body. They are small chains of amino acids synthesized and released by neurons to act as biochemical messengers. They trigger a body-wide response to situations in the environment. Neuroscientist Candace B. Pert has argued that these neuropeptides are the “molecules of emotion” that connect mind and body (Pert, 1997).

A philosopher lives in the world of language. From this abstract point of view learning a language is the key to a link between mental and physical worlds. It is a body that acquires a mind so to speak. Learning a language means to create an extensive isomorphic mapping between the two different worlds of mind and body. Isomorphic means two or more systems, objects, or structures share the same form, shape, or structural properties, often with a one-to-one correspondence between their elements.

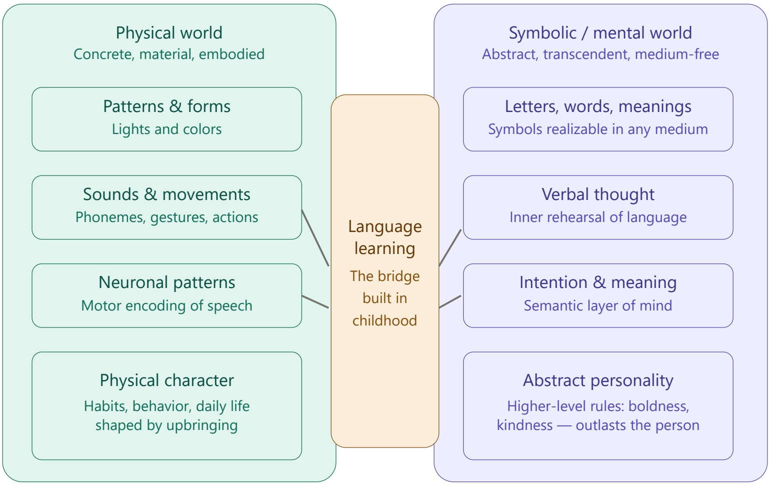
In the opposite direction one could argue that a mind which acquires a body could also create a mind-body connection. For example, a disembodied large-language model (LLM) that learns to operate a robot to create an embodied mind (Varela, et al., 2017).

It is not a dualism of substances but a dualism of worlds. There are clearly two different worlds, not two different substances. There is a mental/symbolic world which consists of letters, words, meanings, intentions, and a physical world which has concrete sounds, movements, physical activity. Letters and words, symbols and tokens, can be realized in any medium, matter or substance. We all learn to bridge the gap when we are young. Learning a language is the link between mental and physical worlds. It means creating an extensive mapping between two different worlds.



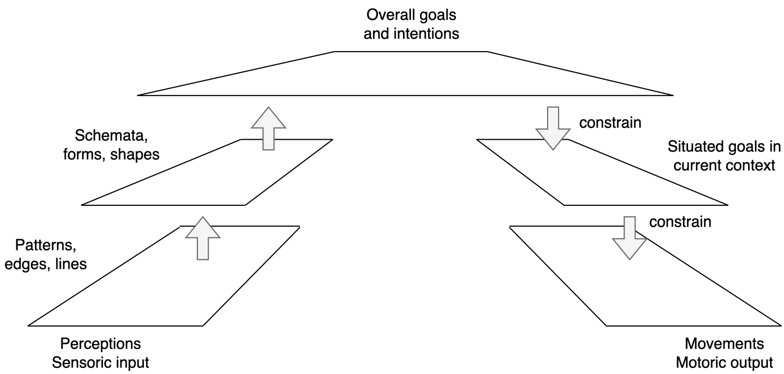
By learning to speak we first link mental entities like letters, words and sentences, to physical objects, like sounds and sequences of vowels or syllables. First we link letters to phonemes and sounds, then words to a sequence of consonants, vowels, and syllables, and finally whole sentences to complex sequences of sounds. To produce the sounds the motoric system needs to encode the corresponding physical actions in neuronal patterns. Additionally, we learn the meaning of a word by mapping it to a physical form or geometric pattern. This is the basic link between abstract symbols and concrete physical actions.

Every word we speak is basically a connection between the abstract world of language, symbols and meaning and the physical word of sounds. It is part of a larger mapping which we understand as language. Every verbal thought we think is a remembrance and rehearsal of a spoken work, and therefore also a connection between the two distinct worlds. We learn to speak, to read and to write for years, and try to establish a mapping of physical and mental properties in a long, painstaking process, but when we are grown-up, we forget all of it and start to ask where the connection between the mind and the body possibly is.



Language alone is not the bridge. A text written in a certain language belongs more to the mental world of ideas. The bridge is created by learning a language, or "in the training of the system" (as we would say for an AI). The process of learning a language is the concrete mechanism that links symbolic thought to physical action. A physical system which has learned a language can be seen as a bridge, because it can operate in both worlds and knows how to connect them.

Can we solve the problem of downward causation too? Is there any or not, and if yes, how does it look like? Downward causation is a causal relationship from higher levels of a system to lower levels of that system. Activity from higher levels of the cortex influence lower levels, for example activity from the prefrontal cortex influences the secondary motor cortex and this in turn influences the primary motor cortex. Overall goals constrain the situated goads in the current context and the spontaneous goals that arise from the current situation, and these intermediate goals in turn constrain concrete actions and movements. This looks like this:



The gap between mental and physical events is bridged on the lowest level where for example the representation of a word is mapped to the sequences of letters to write it or to the sequences of sounds to speak it. We learn to close this gap in school when we learn a language, as we have seen in the first chapter about “intentionality”.

The question how mental events cause physical events seems to be misleading. Mental events primarily do not cause physical events. Mental events cause mental events, and physical events cause physical ones. But the sensory organs in combination with the muscles and the primary somatosensory regions of the brain can provide a bridge between both words. This bridge can be used to connect different agents which both operate in the "mental world" of language. Using this bridge mental events can cause mental events not only in the same agent, but also in other agents. For instance, a person who talks Chinese (to pick one language without loss of generality) sounds like total gibberish to every person who does not speak Chinese. But for another Chinese speakers everything the person says is comprehensible and understandable. The speaker sends out information, and the listener receives and acquires it. Both operate in their "mental world" of Chinese language, which represents the physical world which they both share too.

Levels of Organization

The real question is not if mind and body are two different substances. It is the question how many substances and codes we need if an organizational level is considered as an emergent substance.

Joseph LeDoux defines four nested realms instead of two (mind - body). Instead of using the classic "Mind vs. Body" duality he splits the body into two physical layers and the mind into two mental layers (LeDoux, 2023). The two physical layers are biological and neurobiological, the two mental layers cognitive and conscious:

```
Conscious Realm                  → Subjective awareness, inner experiences
 └── Cognitive Realm             → Mental representations, predictive models
     └── Neurobiological Realm   → Nervous systems, rapid control
         └── Biological Realm    → Metabolism, basic life functions
```

He argues his four "realms" of existence - biological, neurobiological, cognitive and conscious - reflect our evolutionary past (LeDoux, 2023). It starts with plants that are alive, animals that have nervous system, mammals that have large brains, and finally humans that have language and conscious thoughts.

Physical world, nature
↓ DNA
biological system / metabolism / plants
↓ Neural coding
neurobiological system / nervous system / animals

↓ large brains
cognitive system / mammals
↓ Language
conscious system / humans
↓ Written language
Mental world, culture

LeDoux writes we are simultaneously beings on a biological, neurobiological, cognitive and conscious/mental level. LeDoux calls this an "ensemble of being" (LeDoux, 2023).

At the lowest level we have DNA code which uses a four-letter alphabet Adenine (A), Thymine (T), Cytosine (C), and Guanine (G) and is the foundation of every living being. There is the genetic code for the 20 standard amino acids. We have neural coding in the nervous system, the chemical code of neurotransmitters and a neuropeptide code to encode itch and pain signals. We have the alphabet of spoken language, the rules (syntax and semantics) of spoken language and a set of rules of written language.

From a 3rd person point of view the mind can be described as the manifold of all behavior patterns that are typical for a person - essentially an abstract representation of the character or personality. It is in principle a mathematical object if we encode it properly. As an abstract entity it is different from the physical body. It is based on rules and behavior patterns. If we define the mind as the geometric or mathematical "form" of a person's behavioral dynamics, then the mind is a high-level informational pattern rather than a physical substance.

What makes something a "mind" is not what it's made of (carbon or silicon), but how its components are organized and how information flows through the system. In this view mind is an abstract, high-level pattern of causal relationships or behavioral dynamics.

If the mind is an abstract mathematical manifold, it is substrate independent. The abstract pattern can be instantiated in a biological substance, in the neural dynamics of a living body. The physical instantiation is of course part of the body. It is noisy and imperfect. The overall abstract organizational pattern is independent from it. It is the ideal image or type that is never perfectly reached by a physical instantiation. Several factors can prevent the physical body from successfully expressing or reaching its ideal pattern: physical handicaps, a lack of training or learning, diseases, etc.

Just like a “symphony” is an abstract organizational pattern that becomes real if an orchestra plays it - some play it better and some worse - a personality is an organizational pattern that becomes real if a person embodies it.

Instead of having just two levels – mind and body – at least seven core levels of organization can be considered as relevant to understanding how a mind operates: Level 1 – 4 cover the body while level 4 – 7 cover the mind:

Level	Field	Biological Components	Artificial Components
7	Cultural & Social	Laws, Norms, Shared Language, Institutions	AI
6	Psychological / Mind	Beliefs, Desires, Intentions, Thoughts, Traits, Qualia	System Prompt, Steering Vectors
5	Linguistic / Symbolic	Syntax, Semantics, Vector Spaces, Tokens, Concepts	Token Embeddings, Vector Spaces
4	Systems / Networks	Neural Assemblies, Cortical Columns, Circuits	Transformer Attention Heads & Layers
3	Cellular / Biological	Neurons, Synapses, Glial cells	Matrix Mult / Floating-Point Ops (FLOPs)
2	Chemical / Molecular	Amino Acids, DNA, Neurotransmitters, Receptors	TPU / GPU
1	Physical / Quantum	Atoms, Thermodynamics, Electrical Potentials	Electrons & Voltage Transients

For AIs TPUs (Tensor Processing Units) and GPUs (Graphics Processing Units) fulfill the role of neurons in biological systems. For the top levels patterns on level 5 or 6 it does not matter how levels 2 and 3 are implemented.

To come back to the original question: is there a ghost in the machine or not? Yes, but it is not a ghost that feels or thinks. It describes rather the overall, overarching organization pattern of thinking itself. It is the specific high-dimensional geometry of a "personality matrix".

In philosophy it would be a mereological fallacy to say the ghost in the machine (which is part of the machine) does the thinking. A mereological fallacy is logical error of ascribing a property of a whole organism or system to just one of its parts. Bennett and Hacker write in their book “Philosophical Foundations of Neuroscience” which tries to bridge Neuroscience and Philosophy (Bennett, et al., 2021):

> “The principle that psychological predicates which apply only to human beings (or other animals) as wholes cannot intelligibly be applied to their parts, such as the brain, we shall call 'the mereological principle' in neuroscience. Human beings, but not their brains, can be said to be thoughtful or thoughtless; animals, but not their brains, let alone the hemispheres of their brains, can be said to see or not to see, to hear or not to hear, to smell or not to smell and to taste or not to taste things; people, but not their brains, can be said to be decisive or to be indecisive.”

They quote Wittgenstein and his "Philosophical Investigations", where Wittgenstein makes the remark that

> “Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious.”

If we see a ghost as a specific high-dimensional geometry of a “personality matrix”, then clearly both humans and AIs can have a ghost in the machine:

- In humans, personality is encoded in the structural matrix of synaptic connections and chemical balances.
- In AIs based on LLMs, personality is encoded in the structural matrix of the weights in the large-language model (LLM)

The ghost is just a specific high-dimensional geometry of the matrix.

Core References

+ On the Soul, Aristotle
+ Nicomachean Ethics, Aristotle
+ Jaegwon Kim, Philosophy of Mind, Westview Press, 1996
+ Jaegwon Kim, Physicalism, or something near enough, Princeton University Press, 2005
+ Mind in a Physical World, Jaegwon Kim, 1998
+ The concept of Mind, Gilbert Ryle, University of Chicago Press, 1949
+ Being and Nothingness, Jean-Paul Sartre, Simon & Schuster, 1943
+ John Heil, The Universe As We Find It, Oxford University Press, 2015
+ Hannah Critchlow, The Science of Fate: Why Your Future is More Predictable Than You Think, Hodder & Stoughton, 2019

Conclusion

> Therefore, the task is not so much to see what no one has yet seen,
> but to think what no one has yet thought about what everyone sees
> ~ Arthur Schopenhauer (1788-1860)

We have tried to combine the ordinary language philosophy from Gilbert Ryle and his student Daniel Dennett, the philosophy of mind from Peter Strawson and his student John Searle, the pragmatism from Charles Peirce and John Dewey, and the new realism from William James and his student Edwin Holt to investigate the big questions and the unsolved problems in the philosophy of mind.

We followed the tracks of Robert Nozick and his original approaches to see where the solution to the two biggest problems in the philosophy of mind – free will and subjective experience (aka the hard problem of consciousness) – can possibly be found.

The solution to the hard problem of consciousness can be found at an unexpected, unusual place, in the big theaters of show business and film industry, while the solution to the problem of free will sketched in this book follows roughly the original ideas Robert Nozick, Robert Kane, and Alfred R. Mele.

Let us recapitulate the answers to the big questions at "the edge of thought":

- Why are there bad feelings? They are an adaptation to complex, unpredictable environments. Their existence is good for us because they help to avoid that things get worse

- Why do we have subjective experience? For the same reason. Subjective experience evolved because organisms needed rapidly action-guiding evaluations of the world rather than neutral representations

- What is laughter? Laughter is good for us, too. It can be considered as a reward for an insight that is successfully resolving incongruities

- Are we a strange loop? Yes, the self that we recognize in conscious moments of self-awareness has probably the shape of a strange attractor

- Do we have free will? It depends how hard you try. The more you try to have it, the higher the chance you will obtain it. We can try to achieve it, but it is difficult and remains an achievement. Even if the physical universe is entirely deterministic, we still possess a practical, robust form of free will based on conscious, uncoerced deliberation, and we can try to shape our own character by self-forming actions and “time-capsules of intention” for our future selves such as calendar entries, sticky notes and prayers

- Can we solve the hard problem of consciousness? It depends again how hard you try. Since it is a path-dependent problem, we can approximately solve the hard problem by following the path of the subject we try to understand. This leads us to unexpected places: to the big movie theaters of show business and film industry that deliver an answer to the what-it-is-like-to-be question. They do not offer a perfect understanding of subjective experience, but a sufficiently close approximation which is good enough for practical purposes

- Is there a ghost in the machine? Yes, but it is imaginary and elusive because it is tight to the brief conscious moments of self-awareness – unless we consciously develop the personality we want to be, which is hard because rewiring your brain takes time

The hard problems are hard for a reason. As we have seen the hard problem of consciousness can only be solved approximately - by tracking it without completely reaching it, to use the words of Robert Nozick (Nozick, 1981). The problem of free will is hard because it is hard to achieve too. If we are successful in achieving it despite all obstacles, then it is in fact an achievement, as Daniel Dennett said (Dennett, 2003).

It is hard to get rid of the impression that the solutions to the biggest questions can be found in the most ordinary things hiding in plain sight:

- For subjective consciousness humans have built tools to solve the “what it is like to be” question of subjective consciousness. The tools are simply the cinemas and movie theaters which show the stuff that Hollywood produces. They are specifically designed to make the first-person perspective available to a third-person view. Recently the GoPro cameras which show what it is like to be a hero fulfill a similar purpose

- For the question of free will there are well-known and ubiquitous tools too. Not to achieve it, but to take it away: marketing, PR and advertising use all available tricks to manipulate our free will. Would all of them try to take free will away if it does not exist?

- Regarding the afterlife we have built tools too. Churches, mosques, synagogues and temples are places where the magic of religion is used to form groups out of loosely coupled individuals that help us to cope with the big questions of meaning and life after death

One core idea of the book is that exploring unsolved mysteries at the edge of human knowledge requires to leave the abstract ivory tower of a specific discipline, because creative solutions require the interdisciplinary (re-)combination of different perspectives. The solution of the biggest mysteries often hides in plain sight behind the most ordinary objects of daily life. The big, hard and profound problems in philosophy are in fact related to big, existing businesses.

- The hard problem of intentionality is related to learning language in ordinary schools

- The hard problem of free will is related to the big business of PR, marketing and advertising which try to manipulate it

- The hard problem of consciousness is related to nothing else but the show business and film industry in Hollywood and elsewhere

- The hard mind-body problem as the most fundamental problem is related to religious organizations. Some of them operate as "big business”

If we think about it, then it is not surprising that the deep, fundamental questions that everybody cares about have led to big businesses, large organizations and huge corporations. Our social world is made of people. Customers of companies are people. People care about these fundamental questions because they are interested in them and haunted by them, and because so many people care about them, they are obviously a big opportunity to make lots of money.

The problem with philosophy is that philosophers are too much focused-on persons of the past. If you study philosophy or psychology or sociology you are learning the philosophy of Rene Descartes and the philosophy of Immanuel Kant, the psychology of Sigmund Freud and the psychology of Carl Jung, the sociology of Max Weber and the sociology of Karl Marx, etc., without considering all the things we have learned in the meantime. It is true all these thinkers had a distinct lens to view human behavior, but their lens was also distorted by their time. For example, we now know that Marxism did not really work, and we know that there are many things that Descartes simply could not know.

Everybody cites René Descartes (1596-1650) and his dualism although we now know that he was lost in the dark. He lived shortly after the Middle Ages during the Thirty Years' War. What brains are made of has been discovered centuries later. Descartes knew nothing of neurons and synapses and neurotransmitters or of genes, RNA and DNA. He also had no idea that a brain contains 80-100 billion neurons which are linked by trillions of connections. If we rely on his outdated worldview, then we certainly need to ask how the mental is connected to the physical and material, although we now know well that the brain consists of nothing but connections. It contains billions and billions of connections.

The second problem is we tend to forget where we come from and how we grew up. Each of us learns to speak, to read and to write for years, and tries to establish a mapping of physical and mental properties in a long, painstaking process, but when we are grown-up, we forget all of it and start to ask where the connection between the mind and the body possibly is. The connection is formed exactly in these years of childhood when we learn a language.

The third problem is that philosophers apparently do not go to the cinema to enjoy good movies. If they would do it, they would notice sooner or later that they are sitting in a what-it-is-like-to-be machine that shows the audience what it is like to be someone else. And since there is always a lot of advertising in cinemas, they would also notice the free-will-manipulation machine in action. The two biggest problems in the philosophy of mind, the hard problem of consciousness and the problem of free will, would be dissolved in front of their eyes. If there is no free will, why would they try so hard to take it away otherwise? And if there no solution to the hard problem of consciousness, why have we built what-it-is-like-to-be machines that are making lots of money?

And finally, philosophers need to think outside the box and leave their ivory tower of pure thought. Neuroscientists have already demystified many of the hard problems of philosophy. As discussed in the book neuroscience has identified concrete "molecules of emotions", neurotransmitters that connect mind and body, and concrete brain regions which are responsible for meaning and free will, for instance the anterior temporal lobe for meaning and the ventromedial prefrontal cortex for free will.

In other words, the deepest puzzles in philosophy partially appear to be deep because philosophers live in their own, closed world where they cling to the characters and world views of the past and neglect the advances we made in the world we live in. If we look around, we can find convincing answers to all big questions.

I hope this little book was an interesting read, and it contained at least some good points. As Cervantes said, “There is no book so bad...that it does not have something good in it”. Given the current state of our capitalistic and globalized world shortly before a collapse, it would probably have been better to write about the impending catastrophes, the plastic pollution, the limits to growth, the problem of peak oil and climate change, but I feel there is not much what I can do here. Others have written convincingly about it, for example Pablo Servigne and Raphaël Stevens (Servigne, et al., 2020) or Peter Zeihan (Zeihan, 2022) or Donella and Dennis Meadows in their “Limits to Growth” report 50 years ago (Meadows, et al., 1972). Although Dennis has warned about the limits to growth for 50 years, nothing really happened or changed fundamentally. Almost all our politicians are still asking for more growth instead, although it is the path to a collapse of civilization. I would just be another voice which screams in the void, and what is the point of screaming in the void if nobody will listen. If nobody will read it anyway, why not write about something I find interesting myself? At least my future self will be interested in reading it.

Appendix

A. Models for modelling humor

Wilson-Cowan Model combined with Quantum Cognition

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.linalg import expm

# Wilson-Cowan Model combined with Quantum Cognition in Humor
# Simulates three neural assemblies (A, B, C) across a full pre -> post punchline
# trajectory. The punchline occurs at the midpoint of the time axis, activating
# assembly C and shifting the coupling regime.

# Hybrid model which uses the physical firing rates of a Wilson-Cowan model
# to dynamically drive the parameters of a quantum model.
# The top chart shows the physical cause (the biological neurons firing, fighting, and burning energy),
# the normalized neural firing rates of assemblies A, B, C over time
# The bottom chart shows the psychological effect (quantum probabilities in form of the quantum state
# vector reacting to those changing parameters and shifting its phase)
# Both charts share the horizontal time axis.

# --- 1. Simulation Parameters ---
dt = 0.01
t_max = 10.0
steps = int(t_max / dt)
time = np.linspace(0, t_max, steps)

# --- 2. Initialize Wilson-Cowan Variables (Firing Rates) ---
# x_A: Context, x_B: Wrong Meaning, x_C: Punchline
x = np.array([0.2, 0.1, 0.0]) # Initial firing rates
W_inhibit = 1.5             # Baseline inhibition between A and B

# --- 3. Initialize Quantum Variables (Amplitudes) ---
# Start in a state mostly aligned with the initial neural activity
psi = np.array([1.0, 0.0, 0.0], dtype=complex)
psi /= np.linalg.norm(psi)

# History tracking
neural_history = []
quantum_probs = []
metabolic_energy = []

# --- 4. The Unified Hybrid Simulation Loop ---
for i, t in enumerate(time):
  # --- STEP A: Wilson-Cowan Dynamics (The Physical Engine) ---
  # Define inputs that change over time
  if t < 5.0:
    # Joke Setup: Driving both A and B, C is starved of input
    I_ext = np.array([1.2, 1.0, 0.0])
    # A and B inhibit each other
```

```
        dx_A = (-x[0] + 1.0 / (1.0 + np.exp(-(I_ext[0] - W_inhibit * x[1])))) * dt
        dx_B = (-x[1] + 1.0 / (1.0 + np.exp(-(I_ext[1] - W_inhibit * x[0])))) * dt
        dx_C = (-x[2] + 0.0) * dt
    else:
        # Punchline Hits at t=5: Input floods into C, C excites A, C crushes B
        I_ext = np.array([1.0, 0.0, 2.5])
        dx_A = (-x[0] + 1.0 / (1.0 + np.exp(-(I_ext[0] + 1.5 * x[2])))) * dt # C excites A
        dx_B = (-x[1] + 1.0 / (1.0 + np.exp(-(I_ext[1] - 2.5 * x[2])))) * dt # C crushes B
        dx_C = (-x[2] + 1.0 / (1.0 + np.exp(-(I_ext[2])))) * dt

    x += [dx_A, dx_B, dx_C]
    x = np.clip(x, 0, 1) # Keep firing rates bounded between 0 and 1
    neural_history.append(x.copy())

    # Calculate Metabolic Effort (Friction/Tension over time)
    # High firing + high conflict = high metabolic cost
    effort = (x[0] * x[1] * W_inhibit) + 0.1 * np.sum(x**2)
    metabolic_energy.append(effort)

    # --- STEP B: Quantum Dynamics Driven by Neural Layer ---
    # We dynamically construct the Hamiltonian using the current firing rates!
    H = np.array([
        [1.0 - x[0], -W_inhibit * x[1],  1.5 * x[2]],
        [-W_inhibit * x[0], 1.0 - x[1], -2.5 * x[2]],
        [1.5 * x[2], -2.5 * x[2],        2.0 - x[2]]
    ], dtype=complex)

    # Evolve quantum state using the neural-driven Hamiltonian
    U = expm(-1j * H * dt)
    psi = np.dot(U, psi)
    psi /= np.linalg.norm(psi) # Ensure normalization

    # Record quantum probabilities |c|^2
    quantum_probs.append(np.abs(psi)**2)

# --- 5. Plotting the Hybrid Model Results ---
neural_history = np.array(neural_history)
quantum_probs = np.array(quantum_probs)

fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(10, 8), sharex=True)

# Top Plot: Wilson-Cowan Firing Rates & Metabolic Effort
ax1.plot(time, neural_history[:, 0], color='blue', linestyle='solid', label='Neural Pool A (Firing)')
ax1.plot(time, neural_history[:, 1], color='red', linestyle='dotted', label='Neural Pool B (Firing)')
ax1.plot(time, neural_history[:, 2], color='green', linestyle='dashed', label='Neural Pool C (Firing)')
ax1.fill_between(time, metabolic_energy, color='lightgray', alpha=0.15, label='Metabolic Tension (Effort)')
ax1.axvline(x=5, color='black', linestyle=':')
ax1.set_ylabel('Neural Activity / Energy', fontsize=12)
ax1.set_title('Wilson-Cowan Firing Rates Driving Quantum Meaning States', fontsize=14)
ax1.legend(loc='upper right')
ax1.grid(True, alpha=0.2)

# Bottom Plot: Quantum Meaning Probabilities
ax2.plot(time, quantum_probs[:, 0], color='blue', lw=2, label='Meaning A (Context)')
ax2.plot(time, quantum_probs[:, 1], color='red', lw=2, linestyle='dotted', label='Meaning B (Wrong Baseline)')
ax2.plot(time, quantum_probs[:, 2], color='green', lw=2, linestyle='dashed', label='Meaning C (Punchline Interpretation)')
```

```
ax2.axvline(x=5, color='black', linestyle=':')
ax2.set_ylabel('Quantum Probability', fontsize=12)
ax2.set_xlabel('Time (Seconds)', fontsize=12)
ax2.legend(loc='upper right')
ax2.grid(True, alpha=0.2)

plt.tight_layout()
plt.show()
```

B. Models for decision-making and free will

Neural Network Attractor Model using Wilson-Cowan

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.integrate import solve_ivp

# 1. Define Model Parameters
tau = 20.0       # Neural time constant (ms)
W_AA = 4.5       # Self-excitation of Pool A
W_BB = 4.5       # Self-excitation of Pool B
W_AB = 2.5       # Cross-inhibition from B to A
W_BA = 2.5       # Cross-inhibition from A to B
I_base1 = 0.5    # Baseline appeal of Reason 1
I_base2 = 0.5    # Baseline appeal of Reason 2

t_act_of_will = 50 # "act of will"  through deliberation (at t)

def f(s):
   # Sigmoid activation function
   return 1.0 / (1.0 + np.exp(-s))

def act_of_will(t):
   # Simulate Nozick's "act of will" halfway through deliberation (at t = 100ms)
   if t > t_act_of_will:
      I_will1 = 1  # Agent actively assigns weight to Reason 1
      I_will2 = 0
   else:
      I_will1 = 0.0  # Deliberation phase (undecided)
      I_will2 = 0.0

   return I_will1, I_will2

# 2. Define the Dynamical System
def attractor_system(t, y):
   x_A, x_B = y
   I_will1, I_will2 = act_of_will(t)

   # Wilson-Cowan governing differential equations
   dxA_dt = (-x_A + f(W_AA * x_A - W_AB * x_B + I_base1 + I_will1)) / tau
   dxB_dt = (-x_B + f(W_BB * x_B - W_BA * x_A + I_base2 + I_will2)) / tau

   return [dxA_dt, dxB_dt]
```

```
def vector_field(t1 = 1.5, t2 = 0):
    # Calculate vector field at t=150 (after the will has sculpted the landscape)
    U = (-X + f(W_AA * X - W_AB * Y + I_base1 + t1)) / tau
    V = (-Y + f(W_BB * Y - W_BA * X + I_base2 + t2)) / tau
    return U,V

# 3. Run the Simulation
t_span = (0, t_act_of_will)            # Time span in ms
t_eval = np.linspace(0, t_act_of_will, 800)
initial_state = [0.15, 0.14]  # Starting near the unstable saddle point (undecided)
sol = solve_ivp(attractor_system, t_span, initial_state, t_eval=t_eval)

t_span = (0, 250)            # Time span in ms
t_eval = np.linspace(0, 250, 800)
initial_state = [0.15, 0.14]  # Starting near the unstable saddle point (undecided)
sol_complete = solve_ivp(attractor_system, t_span, initial_state, t_eval=t_eval)

# 4. Plotting the Phase Space and Trajectory
fig = plt.figure(figsize=(12, 12))

# Plot 1: Neural Activity Over Time
plt.subplot(2, 2, 1)
plt.plot(sol.t, sol.y[0], label='Pool A (Reason 1)', color='crimson', lw=2.5)
plt.plot(sol.t, sol.y[1], label='Pool B (Reason 2)', color='dodgerblue', lw=2.5)
plt.axvline(x=t_act_of_will, color='gray', linestyle='--', label='Act of Will ($I_{will1}$ injected)')
plt.title('Neural Activity Before Decision Making')
plt.xlabel('Time (ms)')
plt.ylabel('Firing Rate / Activity')
plt.grid(True, alpha=0.3)
plt.legend()

# Plot 1: Neural Activity Over Time
plt.subplot(2, 2, 2)
plt.plot(sol_complete.t, sol_complete.y[0], label='Pool A (Reason 1)', color='crimson', lw=2.5)
plt.plot(sol_complete.t, sol_complete.y[1], label='Pool B (Reason 2)', color='dodgerblue', lw=2.5)
plt.axvline(x=t_act_of_will, color='gray', linestyle='--', label='Act of Will ($I_{will1}$ injected)')
plt.title('Neural Activity After Decision Making')
plt.xlabel('Time (ms)')
plt.ylabel('Firing Rate / Activity')
plt.grid(True, alpha=0.3)
plt.legend()

# Plot 2: Phase Space Vector Field and Trajectory
plt.subplot(2, 2, 3)
X, Y = np.meshgrid(np.linspace(0, 1, 20), np.linspace(0, 1, 20))

# Calculate vector field at t=150 (after the will has sculpted the landscape)
I_will1_post = 0
I_will2_post = 0
U,V = vector_field(I_will1_post, I_will2_post)

stream = plt.streamplot(X, Y, U, V, color='gray', density=1.0)
stream.lines.set_alpha(0.4)
stream.arrows.set_alpha(0.4)
plt.plot(sol.y[0], sol.y[1], color='gray', lw=3, label='Decision Trajectory')
```

```
plt.scatter(initial_state[0], initial_state[1], color='black', zorder=5, label='Initial Indecision')
plt.scatter(sol.y[0][-1], sol.y[1][-1], color='red', marker='X', s=150, zorder=5, label='Decision')

plt.title('Phase Space & Attractor Flow')
plt.xlabel('Activity of Pool A ($x_A$)')
plt.ylabel('Activity of Pool B ($x_B$)')
plt.xlim(0, 1)
plt.ylim(0, 1)
plt.grid(True, alpha=0.3)
plt.legend()

# Plot 3: Phase Space Vector Field and Trajectory
plt.subplot(2, 2, 4)
X, Y = np.meshgrid(np.linspace(0, 1, 20), np.linspace(0, 1, 20))

# Calculate vector field at t=150 (after the will has sculpted the landscape)
I_will1_post = 1.5
I_will2_post = 0
U,V = vector_field(I_will1_post, I_will2_post)

stream = plt.streamplot(X, Y, U, V, color='gray', density=1.0)
stream.lines.set_alpha(0.4)
stream.arrows.set_alpha(0.4)
plt.plot(sol_complete.y[0], sol_complete.y[1], color='gray', lw=3, label='Decision Trajectory')
plt.scatter(initial_state[0], initial_state[1], color='black', zorder=5, label='Initial Indecision')
plt.scatter(sol.y[0][-1], sol.y[1][-1], color='red', marker='X', s=150, zorder=5, label='Decision')
plt.scatter(sol_complete.y[0][-1], sol_complete.y[1][-1], color='black', marker='X', s=150, zorder=5, label='Attractor A
(Chosen)')

plt.title('Phase Space & Attractor Flow')
plt.xlabel('Activity of Pool A ($x_A$)')
plt.ylabel('Activity of Pool B ($x_B$)')
plt.xlim(0, 1)
plt.ylim(0, 1)
plt.grid(True, alpha=0.3)
plt.legend()

plt.tight_layout()
plt.show()
```

C. Paradoxes

The **Agency Paradox** related to free will: who is in control of our actions? If our choices are determined and caused by previous events, we are not in control; but if our choices are not caused by anything, they are random, and we are also not in control. Our choices feel completely personal and independent, and yet they are shaped, enabled, and limited by the social structures, relationships, and environments around us.

The **Perception Paradox**: we think we see the world "as it is" and experience the external world directly, but in reality we see a world distorted by our personal preferences, biases and prejudices. Our minds construct their subjective experiences based on internal filters, personal assessments and expectations.

The **Empathy Paradox** related to subjective experience and the hard problem of consciousness: we can feel another's pain, yet we can never truly leave our own body or enter another's.

The **Ghost in the Machine Paradox**: we are made of mindless matter, yet we are haunted by a mind. We seem to have a mind that is different from the body, and yet we don't have it because we are made of physical, mindless matter. Is there a ghost in the machine or not? How can a physical system contain a separate, non-physical entity as a mind, a soul or a ghost? How can a mind emerge from mindless matter?

The **Identity Paradox**: we are like a river that is constantly changing, yet we believe we are the same person as yesterday.

The **Intentionality Paradox**: mental states can be "about" imaginary objects that do not exist. How can a real mental state inside our heads be about external objects outside the head that are either far away or imaginary?

The **Bad Feelings Paradox**: bad feelings can be good for us. Why?

Index

T

V

W

Bibliography

Arthur, W. Brian. 1994. *Increasing Returns and Path Dependence in the Economy.* University of Michigan Press, 1994.

Barabási, Abert-László. 2016. *Network Science.* Cambridge University Press, 2016.

Barrett, Lisa Feldman. 2017. *How Emotions Are Made.* Pan Macmillan, 2017.

Barrett, Lisa Feldman, Lewis, Michael and Haviland-Jones (Ed), Jeannette M. 2016. *Handbook of Emotions, Fourth Edition.* The Guilford Press, 2016.

Baumeister, Roy F. 2024. *The Science of Free Will: Bridging Theory and Positive Psychology.* Oxford University Press, 2024.

Bennett, Maxwell R. and Hacker, P.M.S. 2021. *Philosophical Foundations of Neuroscience, 2nd edition.* Wiley-Blackwell, 2021.

Bourdieu, Pierre. 1987. *Distinction: A Social Critique of the Judgement of Taste.* Harvard University Press, 1987.

Busemeyer, Jerome R. and Bruza, Peter D. 2012. *Quantum Models of Cognition and Decision.* Cambridge University Press, 2012.

Camus, Albert. 1955. *The Myth of Sisyphus.* Hamish Hamilton, 1955.

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press, 1996.

Chen, Guanrong, Kuznetsov, Nikolay V. and Wang, Xiong. 2021. *Chaotic Systems with Multistability and Hidden Attractors.* Springer, 2021.

Cialdini, Robert B. 2021. *Influence: Science and Practice.* Harper Business, 2021.

Clark, Andy. 2015. *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford University Press, 2015.

Connors, Barry W., Bear, Mark F. and Paradiso, Michael. 1995. *Neuroscience: Exploring the Brain.* Jones and Bartlett Publishers, Inc., 1995.

Crick, Francis. 1994. *Astonishing Hypothesis: The Scientific Search for the Soul.* Scribner, 1994.

Critchlow, Hannah. 2019. *The Science of Fate: Why Your Future is More Predictable Than You Think.* Hodder & Stoughton, 2019.

Damasio, Anonio. 2000. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* Mariner Books, 2000.

Damasio, Antonio. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain.* Vintage, 1994.

—. 2010. *Self comes to mind.* Pantheon Books, 2010.

Dayan, Peter and Abbott, Laurence F. 2001. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* The MIT Press, 2001.
Dennet, Daniel C. 1984. *Elbow Room - The Varieties of Free Will Worth Wanting.* The MIT Press, 1984.
Dennett, Daniel C. 2006. *Breaking the spell.* Viking Books, 2006.
—. 1991. *Consciousness Explained.* Little, Brown and Co., 1991.
—. 1995. *Darwin's Dangerous Idea.* Simon & Schuster, 1995.
—. 2003. *Freedom Evolves.* Viking Books, 2003.
—. 2023. *I've been thinking.* W.W. Norton & Company, Inc., 2023.
—. 1987. *The Intentional Stance.* The MIT Press, 1987.
Dewey, John. 1922. *Human Nature and Conduct.* [https://www.gutenberg.org/files/41386/41386-h/41386-h.htm] Henry Holt and Company, Henry Holt and Company, 1922.
Eagleton, Terry. 2019. *Humour.* Yale University Press, 2019.
Edelman, Gerald Maurice and Tononi, Giulio. 2000. *A Universe of Consciousness.* Basic Books, 2000.
Falk, Emily. 2025. *What We Value: The Neuroscience of Choice and Change.* Profile Books, 2025.
Feinberg, Todd E. and Mallatt, Jon M. 2025. *Consciousness Demystified.* The MIT Press, 2025.
Feldman, David P. 2019. *Chaos and Dynamical Systems.* Princeton University Press, 2019.
Fernyhough, Charles. 2016. *The Voices Within: The History and Science of How We Talk to Ourselves.* Basic Books, 2016.
Fischer, John Martin. 2007. *Four Views on Free Will.* John Wiley & Sons, 2007.
Fischer, John Martin, et al. 2007. *Four Views on Free Will.* Wiley-Blackwell, 2007.
Frank, Robert H. 1988. Passions within Reason – The Strategic Role of the Emotions. W.W. Norton & Company, 1988.
Frankfurt, Harry G. 1998. *The importance of what we care about: philosophical essays.* Cambridge University Press, 1998.
Friedman, Avner and Kao, Chiu-Yen. 2014. *Mathematical Modeling of Biological Processes.* Springer, 2014.
Frijda, Nico. 1986. *The emotions.* Cambridge University Press, 1986.
Fromm, Erich. 1941. *Escape from Freedom.* Farrar & Rinehart, 1941.
Gazzaniga, Michael S., Mangun, George R. and Poeppel, David. 2020. *The Cognitive Neurosciences, 6th edition.* MIT Press, 2020.

Gell-Mann, Murray. 1995. *The quark and the jaguar.* St. Martin's Publishing Group, 1995.
Goffman, Erving. 1963. *Stigma: Notes on the Management of Spoiled Identity.* Prentice-Hall, 1963.
—. 1956. *The Presentation of Self in Everyday Life.* Doubleday, 1956.
Griffith, Meghan. 2021. *Free Will: The Basics (2nd edition).* Routledge, 2021.
Gruner, Charles R. 1978. Understanding Laughter: The Workings of Wit and Humor. Nelson-Hall, 1978.
Haggard, Patrick and Eitam, Baruch. 2015. *The Sense of Agency.* Oxford University Press, 2015.
Haidt, Jonathan. 2003. The moral emotions. [book auth.] K. R. Scherer, & H. H. Goldsmith (Eds.) R. J. Davidson. *Handbook of affective sciences.* Oxford University Press, 2003.
Harris, Sam. 2012. *Free Will.* Free Press, 2012.
Hastings, Alan and Powell, Thomas. 1991. Chaos in a Three-Species Food Chain. *Ecology.* 1991, Vol. 72, 3.
Hofstadter, Douglas. 2007. *I Am a Strange Loop.* Basic Books, 2007.
Holt, Edwin B. 1914. *Concept of Consciousness.* George Allen & Company, 1914.
Holt, Edwin B., et al. 1925. *The new realism.* The Macmillan Company, 1925.
Honderich, Ted, [ed.]. 1995. *The Oxford Companion to philosophy.* Oxford University Press, 1995.
Humphrey, Nicholas. 2022. *Sentience: The Invention of Consciousness.* Oxford University Press, 2022.
—. 2012. *Soul Dust: The Magic of Consciousness.* Princeton University Press, 2012.
James, William. 1918. *The Principles of Psychology, Vol 1.* [https://www.gutenberg.org/files/57628/57628-h/57628-h.htm] Henry Holt and Company, 1918.
—. 1918. *The Principles of Psychology, Vol. 2.* [https://www.gutenberg.org/files/57634/57634-h/57634-h.htm] Henry Holt and Company, 1918.
Jaynes, Julian. 1976. *The Origin of Consciousness in the Breakdown of the Bicameral Mind.* Houghton Mifflin Company, 1976.
Kandel, Eric R., Schwartz, James H. and Jessell, Thomas M. 1981. *Principles of Neural Science.* McGraw-Hill, 1981.
Kane, Robert. 2005. *A Contemporary Introduction to Free Will.* Oxford University Press, 2005.

Kane, Robert L. and Sartorio, Carolina. 2022. *Do we have free will? A debate .* Routledge, 2022.
Kane, Robert. 2024. *The Complex Tapestry of Free Will.* Oxford University Press, 2024.
—. 2011. *The Oxford Handbook of Free Will, 2nd edition.* Oxford University Press, 2011.
—. 1996. *The Significance of Free Will.* Oxford University Press, 1996.
Kemmerer, David. 2026. *The Neuroscience of Word Meaning.* Cambridge University Press, 2026.
Khrennikov, Andrei and Haven, Emmanuel. 2013. *Quantum Social Science.* Cambridge University Press, 2013.
Khrennikov, Andrei. 2010. *Ubiquitous Quantum Structure - From Psychology to Finance.* Springer, 2010.
Kim, Jaegwon. 2005. *Physicalism, or something near enough.* : Princeton University Press, 2005.
Kross, Ethan. 2021. *Chatter: The Voice in Our Head, Why It Matters, and How to Harness It.* Crown Publishing Group, 2021.
Lakoff, George and Johnson, Mark. 1980. *Metaphors We Live By.* University of Chicago Press, 1980.
Lambon Ralph, Matthew A., et al. 2017. The neural and computational bases of semantic cognition. *Nature Revies Neuroscience.* 2017, Vol. 18, pp. 42–55.
Lazarus, Richard S. 1991. *Emotion and adaptation.* Oxford University Press, 1991.
LeDoux, Joseph E. 2019. *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains.* Viking, 2019.
—. 1998. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* Simon & Schuster, 1998.
—. 2023. *The Four Realms of Existence - A New Theory of Being Human.* Harvard University Press, 2023.
Levine, Joseph. 2004. *Purple Haze: The Puzzle of Consciousness.* Oxford University Press, 2004.
—. 2018. *Quality and Content: Essays on Consciousness, Representation, and Modality.* Oxford University Press, 2018.
Liane, Gabora and Kirsty, Kitto . 2017. Toward a Quantum Theory of Humor. *Frontiers in Physics, Sec. Interdisciplinary Physics.* 2017, Vols. Volume 4 - 2016.
List, Christian. 2019. *Why Free Will Is Real.* Harvard University Press, 2019.
Lyman, Monty. 2022. *The Painful Truth: The new science of why we hurt and how we can heal.* Penguin, 2022.

Marx, W. David. 2022. *Status and Culture: How Our Desire for Social Rank Creates Taste, Identity, Art, Fashion, and Constant Change.* Viking, 2022.
Maslow, Abraham. 1954. *Motivation and Personality.* Harper & Brothers, 1954.
McGinn, Colin. 1990. *The Problem of Consciousness.* Blackwell, 1990.
Meadows, Donella H., et al. 1972. *The Limits to Growth.* Potomac Associates - Universe Books, 1972.
Mele, Alfred R. 2009. *Effective Intentions: The Power of Conscious Will.* Oxford University Press, 2009.
—. 2022. *Free Will - An Opinionated Guide.* Oxford University Press, 2022.
—. 2014. *Free: Why Science Hasn't Disproved Free Will.* Oxford University Press, 2014.
Minsky, Marvin. 1986. *The Society of Mind.* Simon & Schuster, 1986.
Mitchell, Kevin. 2023. *Free Agents: How Evolution Gave Us Free Will.* Princeton University Press, 2023.
Morreall, John. 2009. *Comic Relief: A Comprehensive Philosophy of Humor.* Wiley, 2009.
—. 1987. *The Philosophy of Laughter and Humor.* State University of New York Press, 1987.
Musser, George. 2023. *Putting Ourselves Back in the Equation: Why Physicists Are Studying Human Consciousness and AI to Unravel the Mysteries of the Universe.* Farrar, Straus and Giroux, 2023.
Nagel, Thomas. 1986. *The View from Nowhere.* Oxford University Press , 1986.
Nesse, Randolph M. 2020. *Good Reasons for Bad Feelings - Insights from the Frontier of Evolutionary Psychiatry.* Penguin, 2020.
Nozick, Robert. 2001. *Invariances - The Structure of the Objective World.* Harvard University Press, 2001.
—. 1981. *Philosophical Explanations.* Harvard University Press, 1981.
—. 1993. *The Nature of Rationality* . Princeton University Press, 1993.
Oatley, Keith, Keltner, Dacher and Jenkins, Jennifer M. 2020. *Understanding Emotions (4th Edition).* Wiley, 2020.
Panksepp, Jaak. 2005. *Affective Neuroscience: The Foundations of Human and Animal Emotions.* Oxford University Press, 2005.
Peirce, Charles S. 1923. *Chance, Love and Logic.* [https://www.gutenberg.org/files/65274/65274-h/65274-h.htm] Harcourt, Brace and Company, 1923.
Pert, Candace B. 1997. *Molecules of Emotion: Why You Feel the Way You Feel.* Scribner, 1997.

Pierson, Paul. 1994. *Politics in Time: History, Institutions, and Social Analysis.* Princeton University Press, 1994.
Pollan, Michael. 2026. *A World Appears: A Journey Into Consciousness.* Allen Lane, 2026.
Pulvermüller, Friedemann. 2003. *The Neuroscience of Language: On Brain Circuits of Words and Serial Order.* Cambridge University Press, 2003.
Ryle, Gilbert. 1949. *The concept of mind.* University Of Chicago Press, 1949.
Sapolsky, Robert. 2023. *Determined: A Science of Life Without Free Will.* Penguin Press, 2023.
Sartre, Jean-Paul. 1943. *Being and Nothingness.* Simon & Schuster, 1943.
Scott, Alwyn C. 2007. *The Nonlinear Universe: Chaos, Emergence, Life.* Springer, 2007.
Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press, 1983.
—. 2004. *Mind - A Brief Introduction.* Oxford University Press, 2004.
—. 1997. *The Mystery of Consciousness.* New York Review Books, 1997.
—. 1992. *The Rediscovery of the Mind.* The MIT Press, 1992.
Servigne, Pablo and Stevens, Raphaël. 2020. *How Everything Can Collapse: A Manual for our Times.* Polity Books, 2020.
Shankar, Ramamurti. 1994. *Principles of Quantum Mechanics.* Plenum Press, 1994.
Simon, Herbert. 1970. *The Sciences of the Artificial.* The MIT Press, 1970.
Sobchack, Vivian. 1992. *The Address of the Eye: A Phenomenology of Film Experience.* Princeton University Press, 1992.
Solms, Mark. 2022. *The Hidden Spring: A Journey to the Source of Consciousness.* Profile Books, 2022.
Strawson, Peter F. 2008. *Freedom and Resentment and Other Essays.* Taylor & Francis, 2008.
—. 1966. *The Bounds of Sense.* Routledge, 1966.
Strogatz, Steven H. 2014. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering.* Westview Press, 2014.
Sunstein, Cass R. 2025. *Manipulation: What It Is, Why It's Bad, and What to Do About It.* Cambridge University Press, 2025.
Thagard, Paul. 2025. *Dreams, Jokes, and Songs: How Brains Build Consciousness.* Oxford University Press, 2025.
Varela, Francisco J., Thompson, Evan and Rosch, Eleanor. 2017. *The Embodied Mind: Cognitive Science and Human Experience.* The MIT Press, 2017.
Wegner, Daniel. 2002. *The Illusion of Conscious Will.* The MIT Press, 2002.

Wild, Barbara, et al. 2003. Neural correlates of laughter and humour. *Brain : a journal of neurology.* 2003, Vol. 126, 10.
Wilson, Carl. 2014. *Let's Talk About Love: Why Other People Have Such Bad Taste.* Bloomsbury Academic, 2014.
Zeihan, Peter. 2022. *The End of the World Is Just the Beginning - Mapping the Collapse of Globalization.* Harper Collins, 2022.