Could a Machine Think - Churchland&Churchland

// Classical AI is unlikely to yield conscious machines; systems that mimic the brain might

Lau's words

- Thought experiments might not tell us that much

Text

Artificial Intelligence Research

On the history of creating consciousness from machine:
Could a machine think? There were many reasons for saying yes

1. One of the earliest and deepest reasons lay in two important results in computational theory. The first was Church's thesis, which states that every effectively computable function is recursively computable.
    - Effectively computable means that there is a "rote" procedure for determining, in finite time, the output of the function for a given input.
    - Recursively computable means more specifically that there is a finite set of operations that can be applied to a given input, and then applied again and again to the successive results of such applications, to yield the function's output in finite time
2. The second important result was Alan M. Turing's demonstration that any recursively computable function can be computed in finite time by a maximally simple sort of symbol-manipulating machine that has come to be called a universal Turing machine.
    - This machine is guided by a set of recursively applicable rules that are sensitive to the identity, order and arrangement of the elementary symbols it encounters as input.

These Two Results

These two results entail something remarkable, namely that a standard digital computer, given only the right program, a large enough memory and sufficient time, can compute any rule-governed input-output function.

- That is, it can display any systematic pattern of responses to the environment whatsoever.

More specifically, these results imply that a suitably programmed symbol-manipulating machine (hereafter, SM machine) should be able to pass the Turing test for conscious intelligence.

There were a few puzzles, of course. For one thing, SM machines were admittedly not very brain-like. Even here, however, the classical approach had a convincing answer.

1. First, the physical material of any SM machine has nothing essential to do with what function it computes. That is fixed by its program.
2. Second, the engineering details of any machine's functional architecture are also irrelevant, since different architectures running quite different programs can still be computing the same input-output function

Accordingly, AI sought to find the input-output {unction characteristic of intelligence and the most efficient of the many possible programs for computing it. The idiosyncratic way in which the brain computes the function just doesn't matter, it was said. This completes the rationale for classical AI and for a positive answer to our title question

Could a Machine Think?

Could a machine think? There were also some arguments for saying no. Through the 1960's interesting negative arguments were relatively rare.

- The objection was occasionally made that thinking was a nonphysical process in an immaterial soul.
    - But such dualistic resistance was neither evolutionarily nor explanatorily plausible. It had a negligible impact on Ai research.

A quite different line of objection was more successful in gaining the Ai community's attention. In 1972 **Hubert L. Dreyfus** published a book that was highly critical of the parade-case simulations of cognitive activity. He argued for their inadequacy as

simulations of genuine cognition, and he pointed to a pattern of failure in these attempts.

- What they were missing, he suggested, was the vast store of inarticulate background knowledge every person possesses and the commonsense capacity for drawing on relevant aspects of that knowledge as changing circumstance demands

Dreyfus *did not deny* the possibility that an artificial physical system of some kind might think, but he was highly critical of the idea that this could be achieved *solely by symbol manipulation* at the hands of recursively applicable rules. ... (goes on to talk about computation time being a limiting factor (clock speeds for instance))

Worries of the sort Dreyfus had raised finally began to take hold here and there even among AI researchers.

At about this time (1980) John Searle (here) authored a new and quite different criticism aimed at the most basic assumption of the classical research program: the idea that the appropriate manipulation of structured symbols by the recursive application of structure-sensitive rules could constitute conscious intelligence.

- Text goes on to describe the Chinese Room argument
  - "In short, the system is supposed to pass the Turing test, while the system itself lacks any genuine understanding of Chinese or real Chinese semantic content"

(...) In his companion piece in this issue, Searle forthrightly lists a number of these critical responses. We think many of them are reasonable, especially those that "bite the bullet" by insisting that, although it is appallingly slow, the overall system of the room-plus-contents does understand Chinese.

We think those are good responses, but not because we think that the room understands Chinese. We agree with Searle that it does not.
Rather they are good responses because they reflect a refusal to accept the crucial third axiom of Searle's argument: "*Syntax by itself is neither constitutive of nor sufficient for semantics.*" Perhaps this axiom is true, but Searle cannot rightly pretend to know that it is.

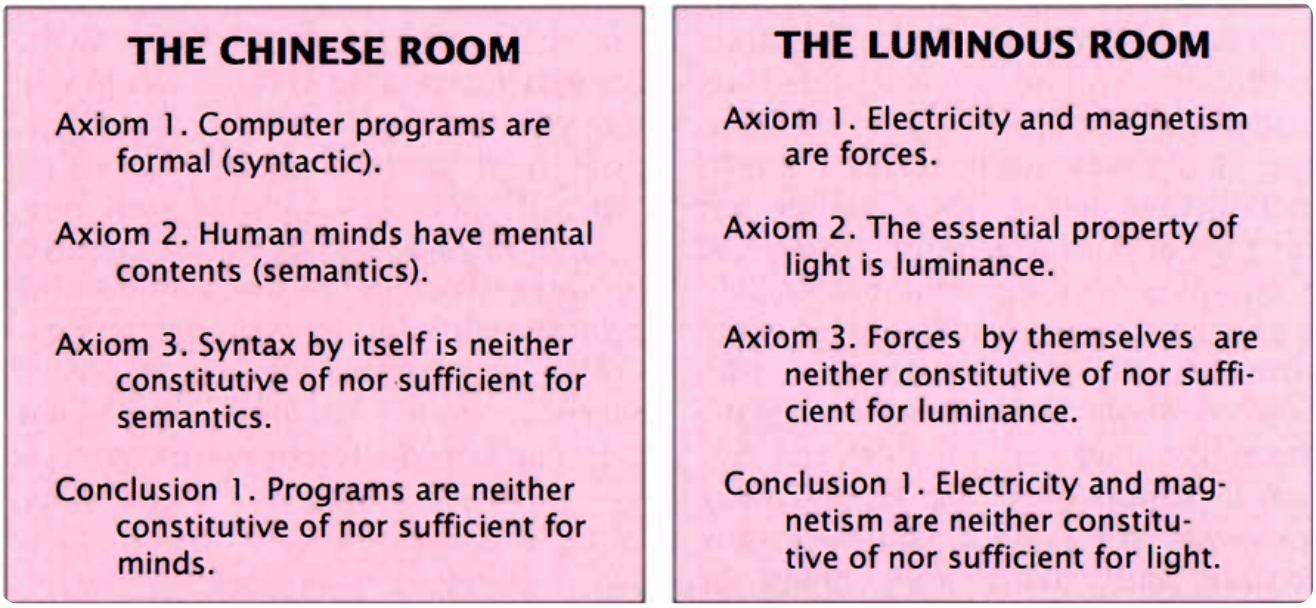
The question-begging character of Searle's axiom 3 becomes clear when it is compared directly with his conclusion 1: "*Programs are neither constitutive of nor sufficient for minds.*"

- Plainly, his third axiom is already carrying 90 percent of the weight of this almost identical conclusion.
- That is why Searle's thought experiment is devoted to shoring up axiom 3 specifically.
    - That is the point of the Chinese room

Although the story of the Chinese room makes axiom 3 tempting to the unwary, we do not think it succeeds in establishing axiom 3, and we offer a parallel argument below in illustration of its failure.

They list a couple of examples of people being wrong about their intuitions of the world...

> "Plainly, what people can or cannot imagine often has nothing to do with what is or is not the case, even where the people involved are highly intelligent"



OSCILLATING ELECTROMAGNETIC FORCES constitute light even though a magnet pumped by a person appears to produce no light whatsoever. Similarly, rule-based symbol manipulation might constitute intelligence even though the rule-based system inside John R. Searle's "Chinese room" appears to lack real understanding.

They then create "The Luminous Room" as a counter argument to that of Searle's Chinese Room. They give the following example as an instance of a counter-argument one could make to the luminous room:

> "Consider a dark room containing a man holding a bar magnet or charged object. If the man pumps the magnet up and down, then, according to Maxwell's theory of artificial luminance (AL), it will initiate a spreading circle of electromagnetic waves and will thus be luminous. But as all of us who have toyed with magnets or charged balls well know, their forces (or any other forces for that matter), even when set in motion, produce no luminance at all. It is inconceivable that you might constitute real luminance just by moving forces around!"

Alas, poor Maxwell has no easy route out of this predicament. All he can do is insist on the following three points.

1. First, axiom 3 of the above argument is false. Indeed, it begs the question despite its intuitive plausibility.
2. Second, the luminous room experiment demonstrates nothing of interest one way or the other about the nature of light.

3. And third, what is needed to settle the problem of light and the possibility of artificial luminance is an ongoing research program to determine whether under the appropriate conditions the behavior of electromagnetic waves does indeed mirror perfectly the behavior of light.

This is also the response that classical AI should give to Searle's argument.

- Even though Searle's Chinese room may appear to be "semantically dark," he is in no position to insist, on the strength of this appearance, that rule-governed symbol manipulation can never constitute semantic phenomena,
    - especially when people have only an uninformed commonsense understanding of the semantic and cognitive phenomena that need to be explained.

==!!we return to the question of whether the research program of classical AI has a realistic chance of solving the problem of conscious intelligence and of producing a machine that thinks.!! ==

- We believe that the prospects are poor, but we rest this opinion on reasons very different from Searle'
- Our reasons derive from the specific performance failures of the classical research program in AI and from a variety of lessons learned from the biological brain and a new class of computational models inspired by its structure.
    - the functional architecture of classical SM machines is simply the wrong architecture for the very demanding jobs required.

What we need to know is this:

How does the brain achieve cognition? Reverse engineering is a common practice in industry.

- In the case of the brain, this strategy presents an unusually stiff challenge, for the brain is the most complicated and sophisticated thing on the planet
- Even so, the neurosciences have revealed much about the brain on a wide variety of structural levels.

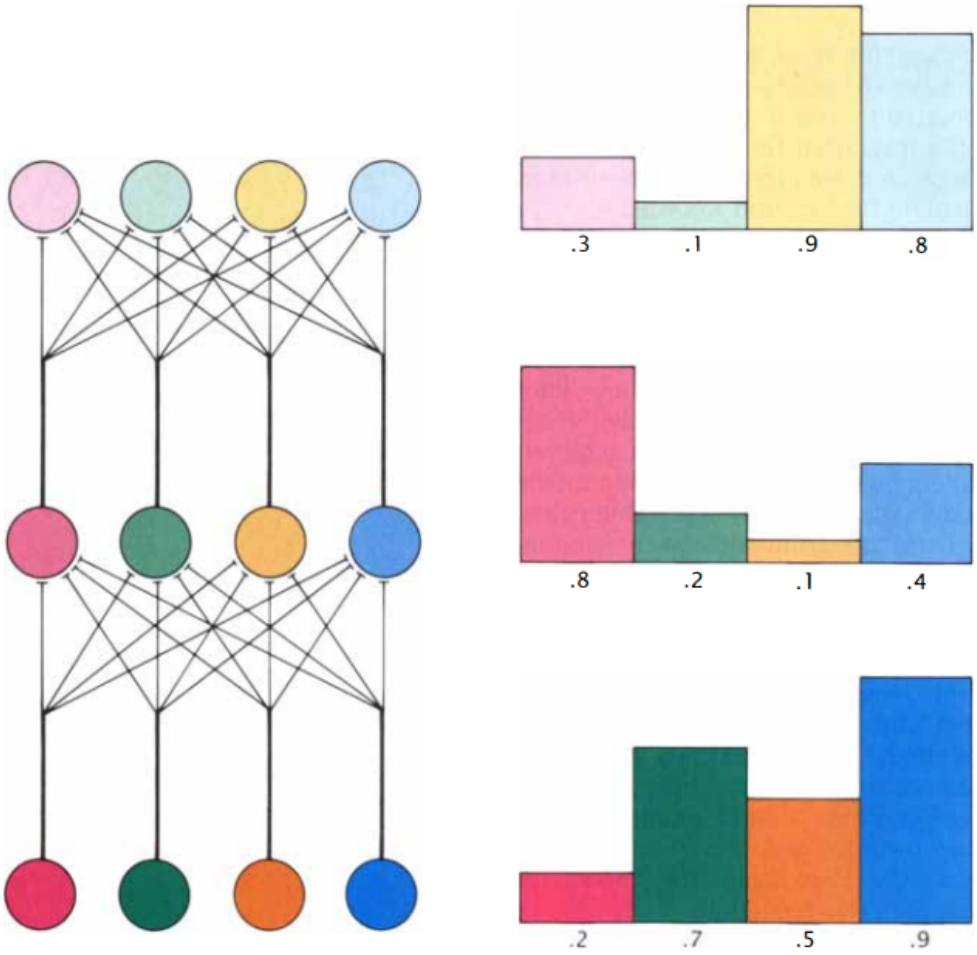
Three anatomic points will provide a basic contrast with the architecture of conventional electronic computers.

1. First, nervous systems are parallel machines, in the sense that signals are processed in millions of different pathways simultaneously.

    - The retina, for example, presents its complex input to the brain not in chunks of eight, 16 or 32 elements, as in a desktop computer, but rather in the form of almost a million distinct signal elements arriving simultaneously at the target of the optic nerve (the lateral geniculate nucleus, (or LGN)) there to be processed collectively, simultaneously and in one fell swoop.
        - see Visual Imagery, Perception, Marr - Vision, A Computational Investigation

2. Second, the brain's basic processing unit, the neuron, is comparatively simple.
    - Furthermore, its response to incoming signals is analog, not digital, inasmuch as its output spiking frequency varies continuously with its input signals.
3. Third, in the brain, axons projecting from one neuronal population to another are often matched by axons returning from their target population.
    - These descending or recurrent projections allow the brain to modulate the character of its sensory processing
    - More important still, their existence makes the brain a genuine dynamical system whose continuing behavior is both highly complex and to some degree independent of its peripheral stimuli.

Highly Simplified model networks have been useful in suggesting how real neural networks might work and in revealing the computational properties of parallel architectures.



NEURAL NETWORKS model a central feature of the brain's microstructure. In this three-layer net, input neurons (*bottom left*) process a pattern of activations (*bottom right*) and pass it along weighted connections to a hidden layer. Elements in the hidden layer sum their many inputs to produce a new pattern of activations. This is passed to the output layer, which performs a further transformation. Overall the network transforms any input pattern into a corresponding output pattern as dictated by the arrangement and strength of the many connections between neurons.

For example, consider a three-layer model consisting of neuronlike units fully connected by axon-like connections to the units at the next layer.

- An input stimulus produces some activation level in a given input unit, which conveys a signal of proportional strength along its "axon" to its many "synaptic" connections to the hidden units.
- The global effect is that a pattern of activations across the set of input units produces a distinct pattern of activations across the set of hidden units.
    - The same story applies to the output units.
- ==All told, this network is a device for transforming any one of a great many possible input vectors (activation patterns) into a uniquely corresponding output vector. ==
    - It is a device for computing a specific function. Exactly which function it computes is fixed by the global configuration of its synaptic weights.

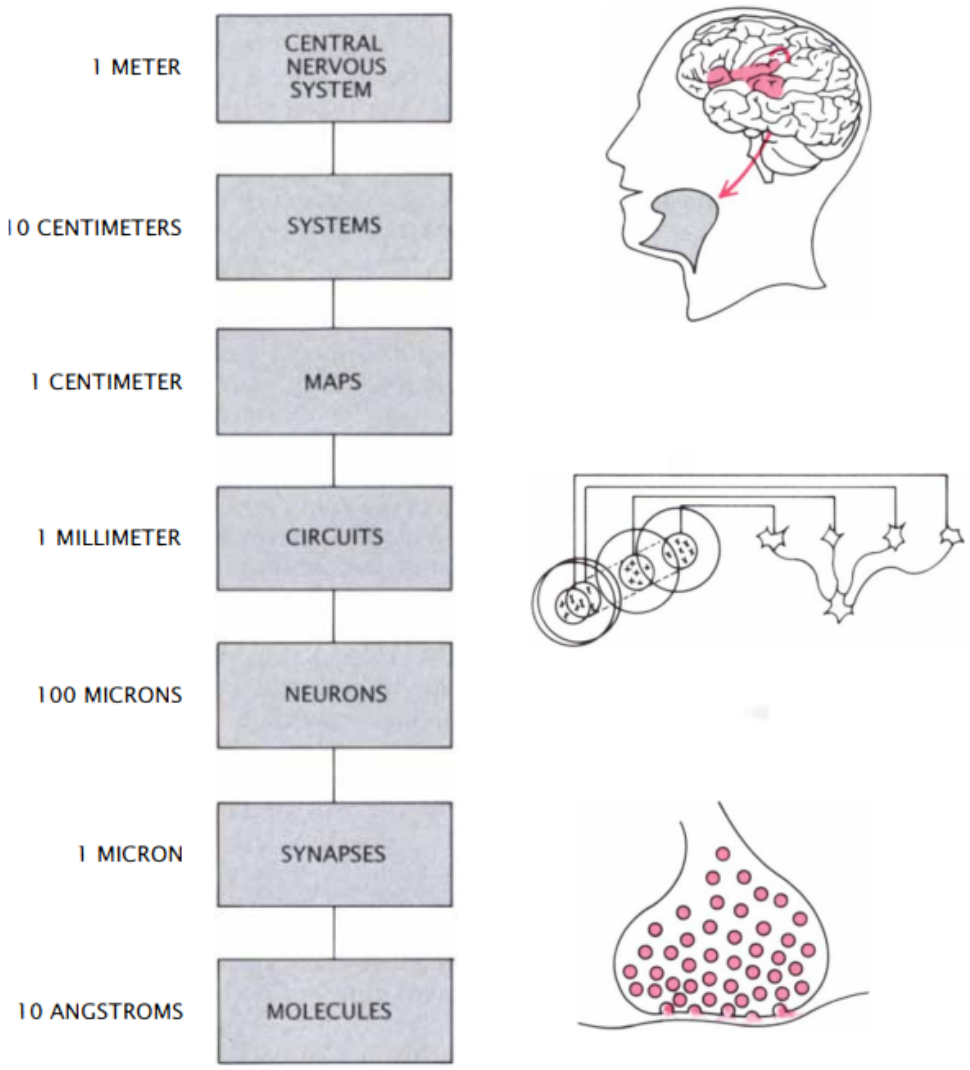
There are various procedures for adjusting the weights so as to yield a network that computes almost any function-that is, any vector-to-vector transformation-that one

might desire.
In fact, one can even impose on it a function one is unable to specify, so long as one can supply a set of examples of the desired input-output pairs.

- This process, called "training up the network," proceeds by successive adjustment of the network's weights until it performs the input· output transformations desired.

It is an oversimplified model of the brain but it does illustrate a couple of important ideas

1. First, a parallel architecture provides a dramatic speed advantage over a conventional computer (parallel-processing)
    - This advantage gets larger as the number of neurons increases at each layer
    - Strikingly, the speed of processing is entirely independent of both the number of units involved in each layer and the complexity of the function they are computing.
2. Second, massive parallelism means that the system is **fault-tolerant and functionally persistent**; the loss of a few connections, even quite a few, has a negligible effect on the character of the overall transformation performed by the surviving network
3. Third, a parallel system stores large amounts of information in a distributed fashion, any part of which can be accessed in milliseconds
    - That information is stored in the specific configuration of synaptic connection strengths, as shaped by past learning.



NERVOUS SYSTEMS span many scales of organization, from neurotransmitter molecules (*bottom*) to the entire brain and spinal cord. Intermediate levels include single neurons and circuits made up of a few neurons, such as those that produce orientation selectivity to a visual stimulus (*middle*), and systems made up of circuits such as those that subserve language (*top right*). Only research can decide how closely an artificial system must mimic the biological one to be capable of intelligence.

Parallel processing is not ideal for all types of computation. On tasks that require only a small input vector, but many millions of swiftly iterated recursive computations, the brain performs very badly, whereas classical SM machines excel (like computing many millions of chess moves in a position very quickly)

- Parallel processing is very convenient for living creatures (like recognizing predators quickly etc.)

Finally, it is important to note that the parallel system described is not manipulating symbols according to structure-sensitive rules. Rather symbol manipulation appears to be just one of many cognitive skills that a network may or may not learn to display. Rule-governed symbol manipulation is not its basic mode of operation.

- Searle's argument is directed against rule-governed SM machines; vector transformers of the kind we describe are therefore not threatened by his Chinese room argument even if it were sound,

- Searle is aware of parallel processors but thinks they too will be devoid of real semantic content.
    - To illustrate their inevitable failure, he outlines a second thought experiment, the Chinese gym, which has a gymnasium full of people organized into a parallel network. From there his argument proceeds as in the Chinese room.

We find this second story far less responsive or compelling than his first. For one, it is irrelevant that no unit in his system understands Chinese, since the same is true of nervous systems: no neuron in my brain understands English, although my whole brain does.

- For another, Searle neglects to mention that his simulation (using one person per neuron, plus a fleet-footed child for each synaptic connection) will require at least 10^{14} people, since the human brain has 10^{11} neurons, each of which averages over 10^{3} connections.
    - It's gonna take a lotta people, many more than there are on earth

On the other hand, if such a system were to be assembled on a suitably cosmic scale, with all its pathways faithfully modeled on the human case, we might then have a large, slow, oddly made but still functional brain on our hands.

- In that case the default assumption is surely that, given proper inputs, it would think, not that it couldn't.
- There is no guarantee that its activity would constitute real thought, because the vector-processing theory sketched above may not be the correct theory of how brains work.
    - But neither is there any a priori guarantee that it could not be thinking.

The brain is a kind of computer,

When brains are said to be computers, it should not be implied that they are serial, digital computers, that they are programmed, that they exhibit the distinction between hardware and software or that they must be symbol manipulators or rule followers.

- Brains are computers in a radically different style.

How the brain manages meaning is still unknown, but it is clear that the problem reaches beyond language use and beyond humans.

- To develop a theory of meaning, more must be known about how neurons code and transform sensory signals, about the neural basis of memory, learning and emotion and about the interaction of these capacities and the motor system.
- A neurally grounded theory of meaning may require revision of the very intuitions that now seem so secure and that are so freely exploited in Searle's arguments.
    - Such revisions are common in the history of science.
        - Kuhn - The Structure of Scientific Revolutions

Could science construct an artificial intelligence by exploiting what is known about the nervous system? We see no principled reason why not.

- Searle appears to agree, although he qualifies his claim by saying that "*any other system capable of causing minds would have to have causal powers (at least) equivalent to those of brains.*"
- We presume that Searle is not claiming that a successful artificial mind must have all the causal powers of the brain, such as the power to smell bad when rotting etc.
- Presumably he means only to require of an artificial mind all of the causal powers relevant, as he says, to conscious intelligence. But which exactly are they?
    - This is an entirely reasonable place for a disagreement, but it is an empirical matter, to be tried and tested.
- Searle hints at various points that every level, including the biochemical, must be represented in any machine that is a candidate for artificial intelligence.
    - This claim is almost surely too strong. An artificial brain might use something other than biochemicals to achieve the same ends

(...) An example is given for an artificially constructed animal retina and cochlear to illustrate the point that bio/neurochemistry is not required for artificial brains

- Whether Mead's program could be sustained to build an entire artificial brain remains to be seen, but there is no evidence now that the absence of biochemicals renders it quixotic

We, and Searle,

> !! "We, and Searle, reject the Turing test as a sufficient condition for conscious intelligence."

At one level our reasons for doing so are similar:

- we agree that it is also very important how the input-output function is achieved; it is important that the right sorts of things be going on inside the artificial machine.

At another level, our reasons are quite different.

- Searle bases his position on commonsense intuitions about the presence or absence of semantic content.
- We base ours on the specific behavioral failures of the classical SM machines and on the specific virtues of machines with a more brain-like architecture.
    - These contrasts show that certain computational strategies have vast and decisive advantages over others where typical cognitive tasks are concerned, advantages that are empirically inescapable.
- But it need not be the only physical system capable of doing so.
    - Artificial intelligence, in a nonbiological but massively parallel machine, remains a compelling and discernible prospect.

Reading Questions

Churchland (P. M.) & Churchland (P. S.) (1990)

1. According to C & C, what is the "fundamental research program of classical AI"?
2. What did classical AI proponents answer to the criticism that symbol-manipulating machines were not very brain-like?
3. Why did the early critiques of classical AI/computationalism from dualists and Dreyfus not initially have much impact on research?
4. What were the reasons that Dreyfus' critique finally took hold?
5. What did Searle presuppose about semantics and consciousness in his Chinese Room argument?
6. What do C & C think about the system reply to the Chinese Room argument?
7. What would the Chinese Room equivalent of (hypothetical) Maxwell's (hypothetical) objection that the Luminous Room is lit to a degree too feeble to appreciate? Does this seem convincing?
8. Which three anatomic points about the brain inspired connectionism?
9. How do C & C define a function? The notion of "impos[ing] on [a network] a function one is unable to specify" may sound a bit like behaviourism's "black box"—but how does it differ?

10. What is C & C's main argument why, allegedly, connectionist networks do not fall prey to Searle's Chinese Room argument?
11. Do C & C claim to have solved the symbol-grounding problem?

Neural Representation and Neural Computation - P. Churchland and Sejnowski

The types of representation and the styles of computation in the brain appear to be very different from the symbolic expressions and logical inferences that are used in sentence-logic models of cognition.

Levels of analysis are not independent in connectionist models, and the dependencies between levels provide an opportunity to co-evolve theories at all levels. This is a radical departure from the *a priori*, introspection-based strategy that has characterized most previous work in epistemology.

1. How do we Represent the World

The central epistemological question, from Plato on is this: *How is representation of a world by a self possible?*

The dominant philosophical tradition has been to try to resolve the epistemological puzzles by invoking mainly intuition and logic...

- Epistemology thus pursued was the product of "pure reason", not empirical investigation, and thus epistemological theories were believed to delimit the necessary conditions, the absolute foundations and the incontrovertible presuppositions of human knowledge

For this a priori task - a task of reflective understanding and pure reason - empirical observations by psychologists and neurobiologists are typically considered irrelevant, or at least, incapable of effecting any significant correction of the a priori conclusion

- Plato, Descartes, and Kant are some of the major historical figures in that tradition; some contemporary figures are Chisholm (1966), Strawson (1966), Davidson (1974), and McGinn (1982).
- It is safe to say that most philosophers still espouse the a priori strategy to some nontrivial extent.

In a recent departure from this venerable tradition of a priori philosophy, some philosophers have argued that epistemology itself must be informed by the

psychological and neurobiological data that bear upon how in fact we represent and model the world.

- First articulated in a systematic and powerful way by Quine (1960)[1], this new "naturalism" has begun to seem more in keeping with evolutionary and biological science and to promise more testable and less speculative answers

There are, undoubtedly, innate dispositions to behave in certain ways and to believe certain things, and to organize data in certain ways, *but innateness is no guarantee of truth*, and it is the truth that a priori reflections are presumed to reveal.

- Innate beliefs and cognitive structures cannot be assumed to be either optimal or true, because all evolution "cares" about is that the internal models enable the species to survive.

It is left for science to care about the truth (or perhaps empirical adequacy)

The a priori insights of the Great Philosophers should be under- stood, therefore, not as *The Absolute Truth* about how the mind- brain must be, but as articulations of the *assumptions* that live deep in our collective *conception* of ourselves.

In addition to asking how the self can know about the external reality, Kant asked: **How is representation of a self by a knowing self possible**?

- One of his important ideas was that the nature of the internal world of the self is *no more unmediated or given than is knowledge of the external world* of physical objects in space and time

A modern version of this insight says: just as the inner thoughts and experiences may represent but not resemble the outer reality, so the inner thoughts may represent but not resemble the inner reality of which they are the representation.

This idea, taken with **Quine's naturalism**, implies that if we want to know how we represent the world-the external world of colored, moving objects, and the internal world of thoughts, consciousness, motives and dreams - the scientific approach is likely to be the most rewarding.

Inner knowledge, like outer knowledge, *is conceptually and theoretically mediated* - it is the result of complex information processing. Whether our intuitive understanding of the nature of our inner world is at all adequate is an empirical question, not an a priori, one

Philosophers-and sometimes psychologists, and occasionally even neuroscientists-generally make one of two responses to the naturalists' conception of the status of our

self-understanding:

1. Philosophy is an a priori discipline, and the fundamental conceptual truths about the nature of the mind, of knowledge, of reason, etc. will come only from a priori investigations (The bounds of science/sense as Strawson (1966) puts it).
    - In a more extreme vein, some existentialist philosophers would claim that the naturalistic approach is itself symptomatic of a civilizational neurosis: the infatuation with science. On this view, the scientific approach to human nature is deeply irrational.
2. Even if a naturalistic approach is useful for some aspects of the nature of knowledge and representation, the neurosciences in particular are largely irrelevant to the enterprise.
    - Neuroscience may be fascinating enough in its own right, but for a variety of reasons it is irrelevant to answering the questions we care about concerning cognition, representation, intelligent behavior, learning, consciousness, and so forth. Psychology and linguistics might actually be useful in informing us about such matters, but neurobiology is just off the book.

2. Why Is Neurobiology Dismissed as Irrelevant to Understanding How the Mind Works?

2.1 The Traditional Problem

In its traditional guise, the mind-body problem can be stated thus: are mental phenomena (experiences, beliefs, desires. etc.) actually phenomena of the physical brain?

- Dualists have answered No to this question.
    - On the dualist's view, mental phenomena inhere in a special, nonphysical substance: the mind (also referred to as the soul or the spirit).
    - The mind, on the dualist's theory, is the ghost in the machine; it is composed not of physical material obeying physical laws, but of soul-stuff, or "spooky" stuff, and it operates according to principles unique to spooky stuff (lol).

The most renowned of the substance dualists are Plato and Descartes, and more recently, J. C. Eccles (1977) and Richard Swinburne (1986). Because dualists believe the mind to be a wholly separate kind of stuff or entity, they expect that it can be understood only in its own terms

- At most, neuroscience can shed light on the *interaction* between mind and body, but not on the nature of the mind itself

It might be thought a bonus of dualism that it implies that to understand the mind we do not have to know much about the brain.

Materialism answers the mind-body question (are mental states actually states of the physical brain?) in the affirmative. The predominant arguments for materialism draw upon *the spectacular failure of dualism to cohere with the rest of ongoing science*.

- Proponents of materialism include Hobbes in the seventeenth century, and in the twentieth, B. F. Skinner (1957, 1976), J. J. C. Smart (1959), W. V. O. Quine (1960), D. C. Dennett (1978) and P. M. Churchland (1984) (husband to author).

Despite the general commitment to materialism, there are significant differences among materialists in addressing the central question of how best to explain psychological states.

- Strict behaviorists, such as Skinner, thought that explanations would take the form of stimulus-response profiles exclusively.
  - Supporting this empirical hypothesis with a philosophical theory, philosophical behaviorists claimed that the mental terminology itself could be analyzed into sheerly physicalistic language about dispositions to behave
  - Curiously perhaps the behaviorists (both empirical and philosophical) share with the dualists the conviction that it is not necessary to understand the workings of the brain in order to explain intelligent behavior

In contrast to behaviorism, **identity theorists** (Smart 1959, Enc 19 claimed that mental states, such as visual perceptions, pains, beliefs, and drives, were in fact identical to states of the brain, though it would of course be up to neuroscience to discover precisely what brain states were in fact identical to what mental states

- The good guys

2.2. The Contemporary Problem: Theory Dualism

Many philosophers who are materialists to the extent that they doubt the existence of soul-stuff nonetheless believe that psychology ought to be essentially autonomous from neuroscience, and that neuroscience will not contribute significantly to our understanding of perception, language use, thinking, problem solving, and (more generally) cognition.

- Thus, the mind-body problem in its contemporary guise is this: Can we get a unified science of the mind- brain? Will psychological theory reduce to neuroscience?

A widespread view (which we call **Theory Dualism**) answers No to the above question. Typically, three sorts of reasons are offered:

1. *Neuroscience is too hard.* The brain is too complex; there are too many neurons and too many connections...
2. *The argument from multiple instantiability*. The argument from multiple instantiability. Psychological states are functional states and, as such, can be implemented (instantiated) in diverse machines (Putnam 1967, Fodor 1975, Pylyshyn 1984).

   - So no *functional* (cognitive) process can be reduced to the behavior of particular neuronal systems (cognition is implementation independent I suppose?)
3. *Psychological states have intentionality*. That is, they are identified in terms of their semantic content; they are "about" other things; they represent things; they have logical relations to one another
   - We can think about objects in their absence, and even of nonexistent objects.
   - "Mars is warmer than Venus" example...
   - In cognitive generalizations states are related semantically and logically, whereas in neurobiological generalizations states can only be *causally* related.
   - Neurobiological explanations cannot be sensitive to the logical relations between the contents of cognitive states, or to meaning or "aboutness".
     - They respond only to causal properties. Neurobiology, therefore, cannot do justice to cognition, and thus no reduction is possible.

2.3. What Is Wrong with Theory Dualism?

In opposition to **theory dualists**, reductionists think we ought to strive for an integration of psychological and neurobiological theory.
Obviously, a crucial element in the discussion concerns what is meant by "reduction"; hence, part of what must first be achieved is a proper account of what sort of business inter-theoretic reduction is.

- Roughly, the account is this: Reductions are explanations of phenomena described by one theory in terms of the phenomena described by a more basic theory.
- Reductions typically involve the co-evolution of theories over time, and as they co-evolve, one theory is normally revised, corrected and modified by its co-evolutionary cohort theory at the other level.
- This revisionary interaction can, and usually does, go both ways; from the more basic to the less basic theory and vice versa
- It is important to emphasize the modification to theories as they co-evolve, because sometimes the modification is radical and entails massive reconfiguration of the very categories used to describe the phenomena.
    - Examples of categories that have undergone varying degrees of revision, from the minor to the radical, include impetus, caloric, gene, neuron, electricity, instinct, life, and very recently, excitability (in neurons) (Schaffner 1976, P. M. Churchland 1979, Hooker 1981).

Because reductionism is frequently misunderstood, it is necessary to be explicit about what is not meant.

- First, seeking reductions of macro-level theory to micro-level theory does not imply that one must first know everything about the elements of the micro theory before research at the macro-level can be usefully undertaken
    - Quite the reverse is advocated-research should proceed at all levels of the system, and co-evolution of theory may enhance progress at all levels
- Additionally, the reduction of theories does not mean that the reduced phenomena somehow disappear or are discredited.
    - The theory of optics was reduced to the theory of electromagnetic radiation, but light itself did not disappear nor did it become disreputable to study light at the macro level
- Whether a category is ultimately rejected or revised depends on its scientific integrity, and that is, of course, determined empirically. (For more detail on inter-theoretic reduction, see P. S. Churchland 1986 (herself))

Given this brief account of reduction as a backdrop, an outline of how the reductionist answers the theory dualist goes as follows:

- Neuroscience *is* hard, but with many new techniques now available, an impressive body of data is available to constrain our theories, and a lot of data are very suggestive as to how neural networks function (see Sejnowski and Churchland, 1987)

    - We have begun to see the shape of neurobiological answers to **functional** questions, such as how information is stored, how networks learn, and how networks of neurons represent.
- *High-level states are multiply instantiable*. So what? (sassy lol).
    - If, in any given species, we can show that particular functional states are identical to specific neuronal configurations (for example, that being in REM sleep is having a specified neuronal state, or that one type of learning involves changing synaptic weights according to a Hebb rule), that will be sufficient to declare a reduction relative to that domain (Richardson 1979, Enc 1983, P. S. Churchland 1986; section 3 below).
    - Moreover, it should be emphasized that the explanation of high- level cognitive phenomena will not be achieved directly in terms of phenomena at the lowest level of nervous-system organization, such as synapses and individual neurons. Rather, the explanation will refer to properties at higher structural levels, such as networks or system.
        - Functional properties of networks and systems will be explained by reference to properties at the next level down, and so on
        - What we envision is a chain of explanations linking higher to next-lower levels, and so on down the ladder of structural levels.
    - The similarity of the information-processing function between two biological systems that are different at the level of the synaptic and cellular levels are probably a consequence of invariants that characterize dynamical systems in high-dimensional state spaces
- *Argument from intentionality*: A theory of how states in a nervous system represent or model the world will need to be set in the context of the evolution and development of nervous systems, and will try to explain the interactive role of neural states in the ongoing neuro- cognitive economy of the system
    - Nervous systems do not represent all aspects of the physical environment; they selectively represent information a species needs, given its environmental niche and its way of life
        - Although the task is difficult, it now seems reasonable to assume that the "aboutness" or "meaningfulness" of representational states is not a spooky relation but a neurobiological relation

Because this answer is highly cryptic and because intentionality has often seemed forever beyond the reach of neurobiology, the next section will focus on intentionality: the theory dualist's motivation, and the reductionist's strategy.

3. Levels, Intentionality, and the Sentence-Logic Model of the Mind

3.1 Sentential Attitudes and the Computer Metaphor

Two deep and interrelated assumptions concerning the nature of cognition drive the third anti-reductionist argument:

Cognition essentially involves representations and computations.

- Representations are, in general, symbolic structures, and
- computations are, in general, rules (such as rules of logic) for manipulating those symbolic structures.

A good model for understanding mind-brain functions is the computer - that is, a machine based on the same logical foundations as a Turing machine and on the von Neumann architecture for a digital computer.

Such machines are ideally suited for the manipulation of symbols according to rules. The computer metaphor suggests that the mind-brain, at the information processing level, can be understood as a kind of digital computer; the problem for cognitive psychology is to determine the program that our brains run.

- The motivating vision here is that cognition is to be modeled largely on language and logical reasoning; having a thought is, functionally speaking, having a sentence in the head, and thinking is, functionally speaking, doing logic, or at least running on procedures very like logic
- ...beliefs, thoughts, hopes, desires and so forth are essential in the explanation of cognition, and that such states are irreducibly semantic because they are identified in virtue of their content sentences
    - That is, such states are always and essentially beliefs that *p*, thoughts that *p*, or desires that *p*, where for *p* we substitute the appropriate sentence
    - Such cognitive states-the so-called **sentential attitudes** - are the states they are in virtue of the sentences that specify what they are about
    - The state transitions are determined by semantic and logical relations between the content sentences, not by casual relations among states neurobiologically described.
    - Thus, cognitive states have meaning (i.e. content, or intentionality), and it might be argued, that it is precisely in virtue of their meaningfulness that

they play the role in cognition that they do

Extending the framework of folk psychology to get an encompassing account of cognition in general, this approach takes it that thinking, problem solving, language use, perception, and so forth will be understood as we determine the sequence of sentences corresponding to the steps in a given information-processing task; *i.e., as we understand the mechanics of sentence crunching*.

- According to this research paradigm, known as **sententialism**, it is the task of cognitive science to figure out what programs the brain runs, and neuroscience can then check these top-down hypotheses against the wetware to see if they are generally possible. (See especially Fodor 1975, Fodor 1981, and Pylyshyn 1984.)

3.2. Is Cognition Mainly Symbol Manipulation in the Language of Thought?

Although this view concerning the nature of cognition and the research strategy for studying cognition may be appealing (much of the appeal is derived from the comfortable place found for folk psychology), it suffers from major defects:

- Many cognitive tasks, such as visual recognition and answering simple true-or-false questions, can be accomplished in about half a second.
  - Given what we know about conduction velocities and synaptic delays in neurons, this allows about 5 milliseconds per computational step, which means that there is time for only about 100 steps.
  - For a sequential program run on a conventional computer, 100 steps is not going to get us remotely close to task completion. Feldman and Ballard (1982) call this the **hundred-step rule.**
- Anatomically and physiologically, the brain is a parallel system, not a sequential von Neumann machine.
  - The neural architecture is highly interconnected. Neurons such as Purkinje cells may have upwards of 80,000 input connections, and neurons in cerebral cortex can have upwards of 10,000 output connections
- However information is stored in nervous systems, it appears to be radically unlike information storage in a digital computer, where storage and processing are separated and items are stored in memory according to addressable locations.
  - In nervous systems, information seems to be stored in the connections between the same neurons that process the information

    - There does not appear to be a distinct storage location for each piece of stored information, and information is content addressable rather than location addressable.
    - Information storage is probably at least somewhat distributed rather than punctate, since memories tend to be degraded with damage to the system rather than selectively wiped out one by one
- A task may fall gracefully on to one architecture, and not on to another. Certain kinds of tasks, such as numerical calculation, fall gracefully on to a von Neumann architecture, but others, such as learning or associative memory, do not
- The hardware-software analogy fails for many reasons, the most prominent of which are that nervous systems are plastic and that neurons continually change as we grow and learn... A von Neumann machine is rigid and fault intolerant
- The analogy between levels of description in a conventional computer (such as the hardware-software distinction) and levels o explanation in nervous systems may well be profoundly misleading.
    - Exactly how many of levels of organization we need to postulate in order to understand nervous-system function is an empirical question, and it may turn out that there are many levels between the molecular and the behavioral.
- Nonverbal animals and infraverbal humans present a major problem for the sentence-logic theory of cognition: How is their cognition accomplished?

If cognition, then, is *not*, in general, to be understood on the sentence-logic model, the pressing questions then are these: How d the brain represent? How do nervous systems model the external world of objects in motion and the internal world of the nervous system itself?

- In certain respects, the current scientific state of a general theory of representation is analogous to the science of embryology in the nineteenth century
    - The development of highly structured, complex, fully-formed organisms from eggs and sperm is a profoundly amazing thing.
    - Faced with this mystery, some scientists concluded that the only way to explain the emergence of a fully structured organism at birth was to assume that the structure was already there
        - Hence the homuncular theory of reproduction, which claimed that a miniature but complete human already exists in the sperm and merely expands during its tenure in the womb (wtf)
        - ...but mainly in the form of DNA-a molecule that looks not at all like a fully formed human. Thus, the structure of the cause does not

resemble the structure of the effect

Instead of starting from the old sentence-logic model, we model information processing in terms of *the trajectory of a complex nonlinear dynamical system in a very high-dimensional space* (just as I said).

- This structure does not resemble sentences arrayed in logical sequences, but it is potentially rich enough and complex enough to yield behavior capable of supporting semantic and logical relationships

We shall now explore what representing looks like in a particular class of nonlinear dynamical systems called connectionist models.

4. Representation in Connectionist Models

Write more stuff here (pg. 14 ish)

Reading Questions

Churchland (P. S.) & Sejnowski (1990):

1. In what ways has epistemology been biased so far, and which negative consequences has this had?
2. Even though most behaviorists would see themselves as materialists (rather than as dualists), which conviction do they share with dualists?
3. In the authors' view, what would it mean if "psychological theory reduce[s] to neuroscience"?
4. Which of the seven "major deficits" of (i.e., criticisms towards) the notion that cognition is symbol manipulation in the Language of Thought do you find the most (and the least) convincing? Why?
5. What do the authors mean when they say that connectionist models "have clear ties with the computational level […] and with the implementation level" (cf. Marr, 1982)? Is that true?
6. What difference did the initial condition set for NETtalk make? If this is how the brain works, what does that tell us?
7. What makes backpropagation "biologically implausible"? Why may it still be a good "system-level algorithm"?
8. How does connectionism wish to revise Marr's model?

9. What are the authors' argument why cognitive science needs to be an interdisciplinary endeavour?

Footnotes

1. Not the same article I believe but same guy ↩

Mock Exam - Connectionism

Representation:
Computationalims: Symbolic
Connectionism: The representation arises from the activity of the network (emergent property of a complex system)

Implementation:
Computationalism: The program could run in anything = implementation independent
Connectionism: Dependent on a network of neurons (such as a brain) or anything else brain-like

Mock Exam Question

Explain how connectionism differs from computationalism, and how it may provide better models for explaining the functions of the human mind.

- (In doing so, describe the components of a neural network and how that differs from a serial computer)
- Discuss how semantics may emerge from a connectionist network and whether that offers a solution to the symbol grounding problem.

Computationalism vs connectionism

- Comp: Purely symbolic (formal syntax/symbol manipulation)
- Conn: Still symbolic but emergent from a complex system
    - Neural network (connectionism):
        - A series of layers which work in parallel
        - Layers have "neurons" or specific weights/activations which are adjusted continually

        - The neuron with the most activity in the output layer (the most probable answer according to the network)
        - Activity learning: the model "learns" through repeated exposure/data feeding
            - Same as human experience --> learn through trial and error
            - Aligns with established neuroscience knowledge: Feature detectors, backpropagation, Hebb's rule ("cells that fire together, wire together")
        - Way better at solving perception related tasks e.g. letter-recognition
        - Implementation: has to run on a network of neurons (such as a brain)
    - A serial computer (computationalism):
        - Works in serial instead of parallel (sequential execution)
        - Does not adjust weights concurrently but runs pre-defined algorithms/programs (doesn't learn on the fly)
        - Implementation: Can run on anything that can represent 0s and 1s (implementation independent)
- How semantics emerge from a connectionist network

  - The semantics arises from the syntax (the relations between weights and activations of the neurons of the trained model) ^a355b0
  - Represents concepts in very high dimensional vector space (each dimension corresponds to a feature/attribute)
  - This representation allows the network to capture complex relationships between contexts and their conceptual/semantic meaning

Connectionism Lecture

Week 15 - Connectionism

- Computation happens below the representational level in a connected network of units (e.g. neurons)
- Semantics emerges from the network –
- The algorithm level is not independent from the implementation level

Learning Goals

1. Understanding the limitations of the computationalists paradigm seen from the connectionist paradigm

2. Being able to reflect on the values of thought experiments
3. Understanding how neural networks work on a broad level
4. Understanding how semantic content may emerge from the network, i.e. how symbols may get grounded

Overview

- Philosophy of science
  - Connectionism: a paradigm shift
- Revisiting thought experiments
  - How much can they tell us?
- Connectionist cognitive science
  - Tensions and shortcomings in computationalism
  - Structure of connectionist networks
  - Cognition is what brain-like things do
    - Cognition is the activity of connected networks
    - i.e. cognition cannot be implemented in any hardware

Paradigms

Kuhn - The Structure of Scientific Revolutions

The practices that define a scientific discipline

- What phenomena do we look at?
- Which questions do we ask?
- How do we design experiments?
- How do we interpret results?

The computationalist paradigm

Computationalist practices

- Range of behaviour made possible by internal computation
- Which computational processes underlie this range of behaviour?
- Test the flexibility and constraints of cognition
- Infer the inner computational operations

Revisiting thought experiments

On thought experiments

We are taking something abstract ... e.g. implementation independent, systematically interpretable symbol manipulation

- ... and making it more vivid and relatable ...

“Plainly, what people can or cannot imagine often has nothing to do with what is or is not the case ...” p. 35 Could a Machine Think - Churchland&Churchland

The Luminous Room Chinese Room vs Luminous Room.png

- What is the luminous room argument supposed to establish?
- What do the Churchlands have to say about the third axiom?
- What is a precondition according to Searle for arguments of analogy to work?
- Does it weaken the Chinese Room argument?

3 reasons against the luminous room argument

1. Axiom 3
   - 

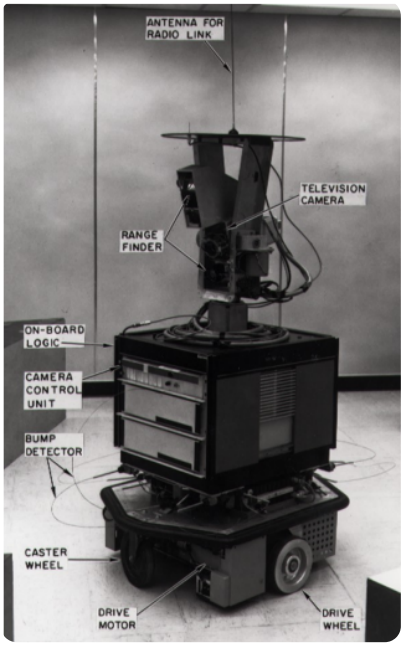
2. our intuitions doesn’t constrain nature
   - “... the luminous room experiment demonstrates nothing of interest one way or the other about the nature of light” p. 35
   - our intuitions are experience-dependent (and culture dependent)
     - Be careful with how much weight you give your intuitions
3. the question is empirical
   - Human Centric.png
   - More objective description.png

Searle's Rebuttal

Analogy fails because: the electromagnetic radiation is a causal story; on the contrary, formal symbols have no causal power

“... what is needed to settle the problem of light and the possibility of artificial luminance is an **ongoing research program** [my emphasis] to determine whether under the appropriate conditions the behavior of electromagnetic waves does indeed mirror perfectly the behavior."

Classical Artificial Intelligence

- SHAKEY
  - 

  - Shakey was a shitty robot apparently
  - Took a couple of minutes to navigate stuff

1. Signal propagation in the computer much faster than in the brain (~$1e^6$ times faster)
2. Clock frequency of the central processing unit is higher by a similar margin
3. Conclusion: we are doing something wrong in modelling perception

The context problem

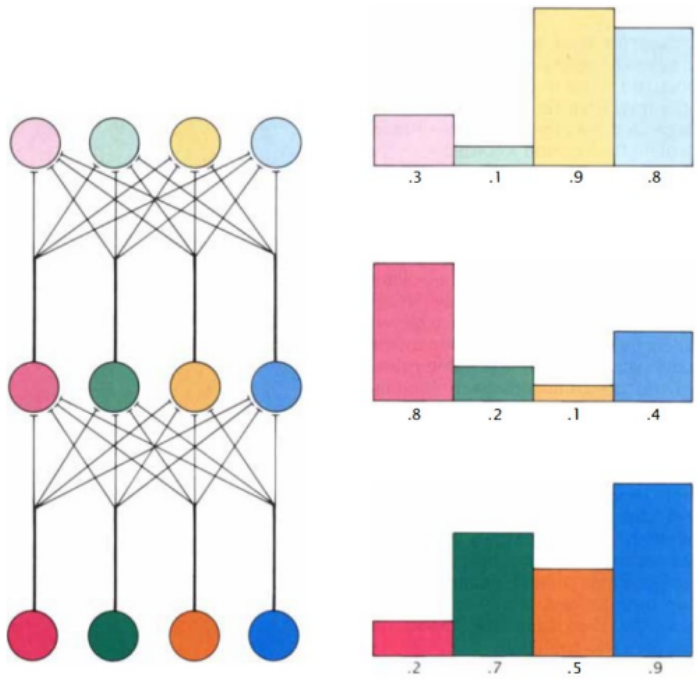
- Humans
  - Have inarticulate background knowledge
  - Which can be applied in a flexible manner to changing contexts
- Classical artificial intelligence
  - Background knowledge needs to be stored explicitly

- Access faces the double problem that:
  - Search time is long
  - And what to search for? (what is the context)

Connectionist cognitive science

- REVERSE ENGINEERING THE BRAIN (used to be the mind but we're hoping that comes with it)

1. Brains are *parallel* machines
2. Neurons, its processing units are *analogue* (in terms of spiking frequency) (can be between 0&1, not just 0&1)
3. Projections are both *feedforward* and *recurrent*



- LEFT: Weights
- RIGHT: Activations

Advantages of parallelism

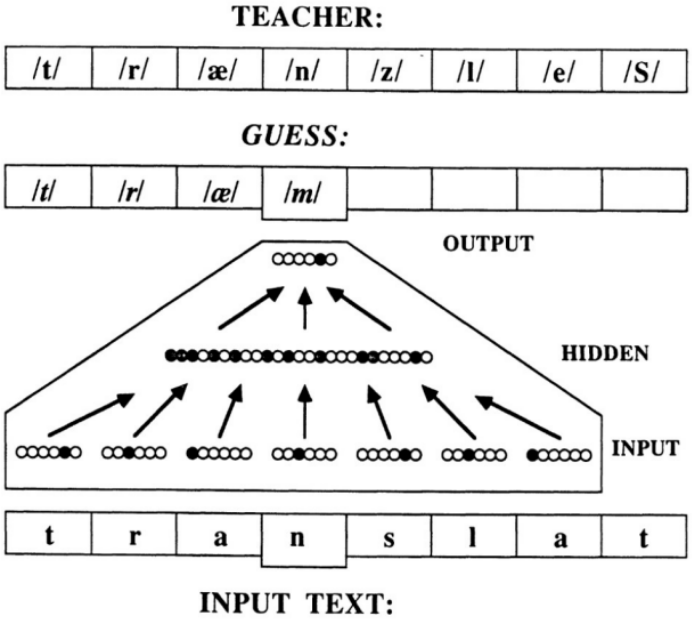
1. Speed
2. Redundancy – fault tolerance
3. Flexible storage

Interim summary

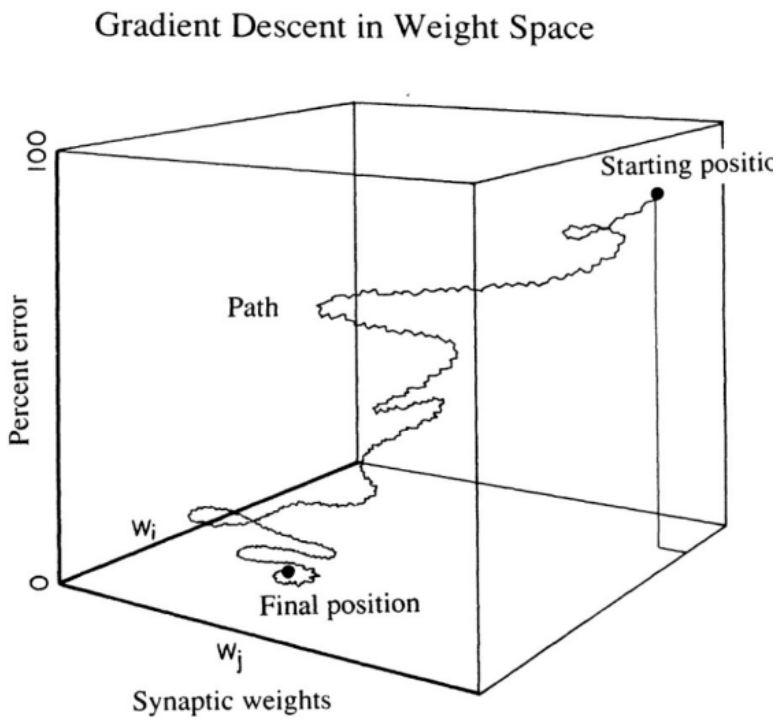
- Connectionism is a (small?) paradigm shift away from computationalism
- The relevance of thought experiments can be called into question
- Serial computationalism faced problems of speed and context
- Connectionist network are more brain-like (implementation independence?)
- (Also check out 3Blue1Brown's videos on the matter: (The goat)

Do psychology/physiology and connectionist networks look alike?

Old examples:

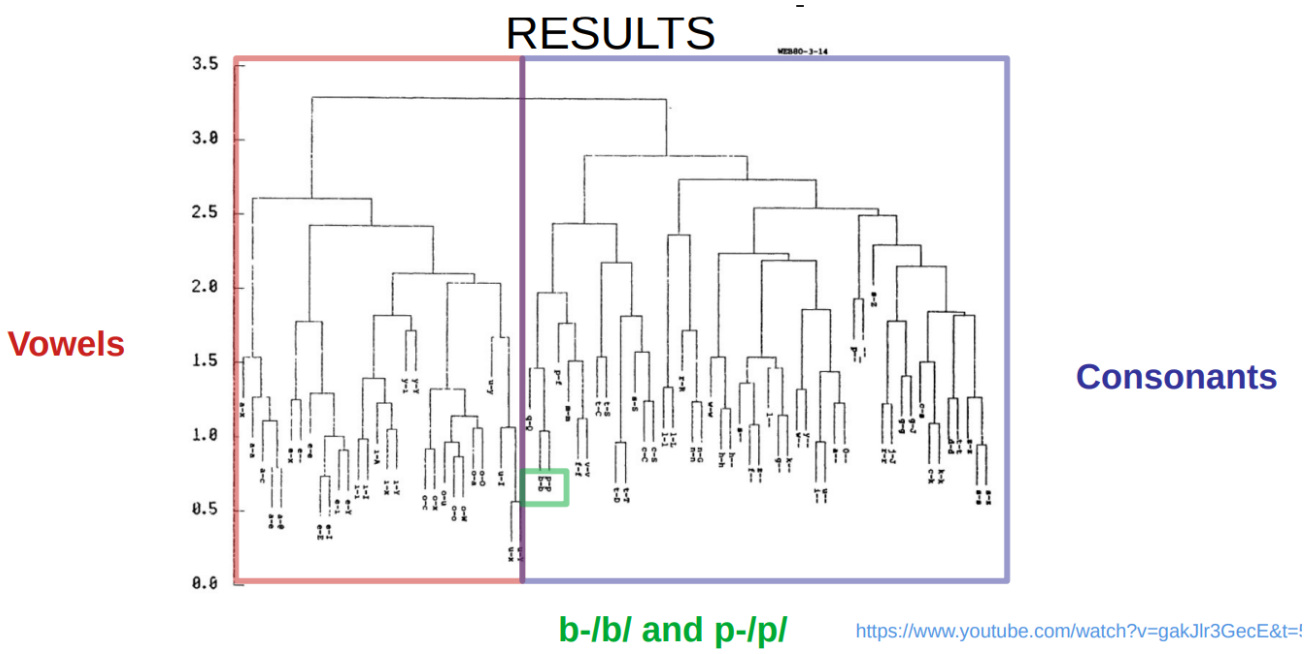


- GRADIENT DESCENT



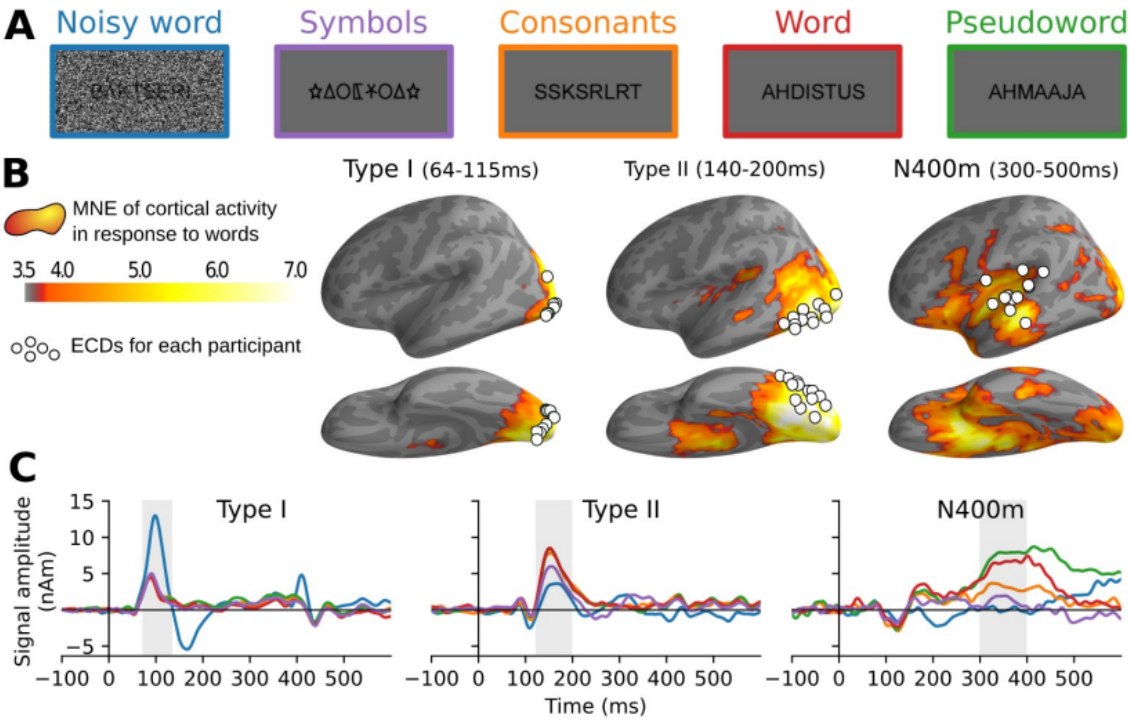
How to optimise gradient descent?
BACKPROPAGATION

- We want to minimize the cost function
- So it is *supervised* learning

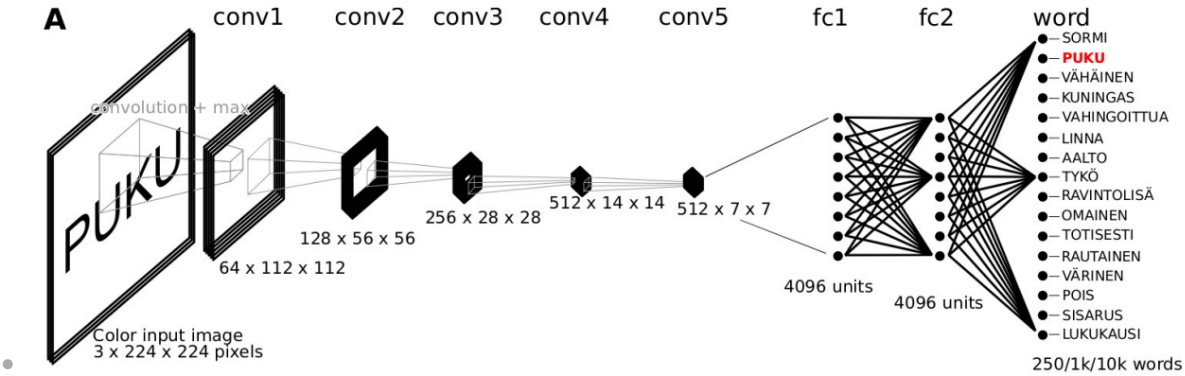


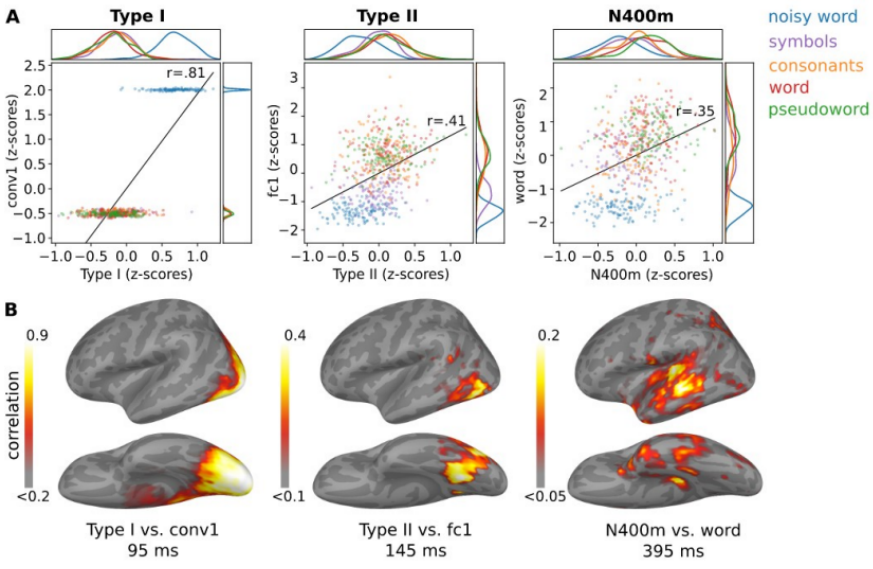
Do brains and connectionist networks look alike?

A modern example



[1]





Can we preserve implementation independence?

- NO! — ALL LEVELS ARE MIXED

Computational theory

- Computational functions and limitations in humans are dependent on “hardware” and what can be represented in such hardware

Hardware implementation

- Brain architecture constrains functions and representations

Representation and algorithm

- Algorithms that cannot be on run on brain hardware not interesting

Representation

- AN ANSWER TO THE SYMBOL GROUNDING PROBLEM?

These two examples suggest that the representation emerges from the structure of the neural network

- The representation is thus not symbolic anymore
- Rather the semantics arises from the syntax (the relations between weights and activations of the neurons of the trained model)
    - I.e. Searle’s Axiom 3 is false

Definition of **emergent properties**:

- *A property of a complex system is said to be ‘emergent’ just in case, although it arises out of the properties and relations characterizing its simpler constituents, it is neither predictable from, nor reducible to, these lower level-characteristics.”* [2]

Searle’s reply to the brain simulator reply

> “Computer simulations of brain processes provide models of the formal aspects of these processes. But [...] the computational model is no more

real than the computational model of any other natural phenomenon” p. 29

External Interpretation

CONNECTIONIST NETWORK GOALS ARE SET EXTERNALLY

What if we equip our connectionist networks with sensorimotor processing (T3)?

- DO THEY GET CLOSER TO GROUNDING THEIR REPRESENTATIONS?
- https://www.youtube.com/watch?v=djzOBZUFzTw&t=158s

Summary

- Connectionist networks can solve tasks, e.g. letter-to-phoneme and symbol-to-word, that classical computationalism struggled with
- Semantics emerges from the connections and activations between neurons
- Marr’s three levels cannot be kept in separation
- Goals need to be set externally
    - May adding sensorimotor processing allow for goals to be set internally?

Footnotes

1. Vliet, M. van, Rinkinen, O., Shimizu, T., Niskanen, A.-M., Devereux, B., Salmelin, R., 2023. Convolutional networks can model the functional modulation of MEG responses during reading. DOI↩
2. Kim, J., 2005. Emergent properties. The Oxford Companion to Philosophy.↩

1 Perception as causal inference

Predictive processing

Abstract

Our senses are bombarded with input from things in the world. On the basis of that input, we perceive what is out there. The problem that is the focus here is how the brain accomplishes this feat of perception.

- This chapter pursues the idea that the brain must use inference to perceive — *The brain is a Bayesian inference mechanism*

1. The first aim is to show why we should agree with this and what the key ingredients of such perceptual inference are.
2. The second aim is to show how inference could underpin the phenomenology of perception.

The chapter describes how perceptual inference is embedded in a perceptual hierarchy of increasing time scales, maintained in the brain.

This gives reason to believe that perceptual inference can accommodate the richness of perceptual phenomenology, as perception encompasses both variant and invariant representation; much of this is illustrated with examples from perceptual science, in particular binocular rivalry.

- The chapter ends with a brief primer on Bayes' rule.

Introduction

A very basic and useful formulation of the problem of perception is in terms of cause and effect. States of affairs in the world have effects on the brain—objects and processes in the world are the causes of the by sensory input. The problem of perception is the problem of using the effects—that is, the sensory data that is all the brain has access to—to figure out the causes.

- It is then a problem of causal inference for the brain, analogous in many respects to our everyday reasoning about cause and effect, and to scientific methods of causal inference.

Different causes can give rise to very similar effects on our sense organs. Consider the potentially identical sensory input from different objects such as a bicycle or a mere picture of a bicycle, or a whole bicycle occluded by a bush as opposed to detached bicycle parts strewn around a bush, or more outré possibilities such as it being an unusually well-coordinated swarm of bees causing the sensory impression as of a bicycle.

In our complex world, there is not a one-one relation between causes and effects, different causes can cause the same kind of effect, and the same cause can cause different kinds of effect.

- This makes it difficult for the brain to pick the one effect (sensory input) that goes with the one cause (object in the world)

Constraints on perceptual inference

The key issue is then that without any additional constraints the brain will not be able to perform reliable causal inference about its sensory input.

- We can in fact engage in such inference, since we can perceive. So there must be such additional constraints, but what could they be?

1. One possibility is that the additional constraints are mere *biases*. Even though the brain cannot reliably infer that it is one rather than another cause, it simply happens to be biased in favor of one. It just so happens that it decides in favor of, say, the bicycle being the cause when it gets a certain kind of input'
    - But even if at some level of description there are these regularities it would not solve the problem of perception as we have conceived it.
    - Such regularities do not afford an understanding of perception as causal inference
        - Inference is a **normative** notion and brute biases cannot lead us to understand how there could be a difference in quality between an inference back to bicycles
    - What brute regularities in nature give us is a story about what the system *would* do, not what it *should* do in order to get the world right.
    - What is needed, then, is a **normative** understanding of the role of such regularities. We need to see the additional constraints on causal inference in normative terms.

There is a clear first candidate for an additional constraint with normative impact. It seems obvious that causal inference about things like bicycles draws on a vast repertoire of *prior belief*.

- On the one hand, if the story we tell is that we just find ourselves with a stock of prior beliefs, then we have not after all moved beyond the mere biases type of story.

- On the other hand, if prior knowledge is itself a product of prior perceptual, causal inference, then we are presupposing what we set out to explain, namely perceptual causal inference—the bump in the carpet has merely shifted.

We can now see what a solution to the problem of perception must do. It must have a **bootstrapping effect**[1] such that perceptual inference and prior belief is explained, and explained as being normative, in one fell swoop, without helping ourselves to the answer by going beyond the perspective of the skull-bound brain (Eliasmith 2000; Eliasmith 2005).

- It is based in probability theory—*Bayesian epistemology*— which is normative because it tells us something about what we should infer, given our evidence.
- The Brain and Decision Making > Bayes' Theorem

Perception and Bayes' rule

Consider this very simple scenario. You are in a house with no windows and no books or internet. You hear a tapping sound and need to figure out what is causing it

- There is no end to the possible causes.
- Call each of these possibilities a *hypothesis*. The problem of perception is how the right hypothesis about the world is shaped and selected.

We are able to appreciate the link between a hypothesis and the effects in question. We can say “if it is really a woodpecker, then it would indeed cause this kind of sound”. We can say something about how likely it is that the hypothesis fits the effects.

- This is **likelihood**: the probability that the causes described in the hypothesis would cause those effects.
  - See this for more on likelihood vs probability: Elliot Sober and Creationism > Paley's Watch and the Likelihood Principle
- It is clear that assessing such likelihoods is based on assumptions of causal regularities in the world (for example, the typical effects of woodpeckers).
- Based on our knowledge of causal regularities in the world we can often rank hypotheses according to their likelihood, according to how close their tie is to the effects we are seeking to explain.
  - Such a ranking can be said to capture how good the hypothesis is at accounting for, or predicting, the effects

Just going by the hypothesis with the very highest likelihood does not ensure good causal inference

- Cunning neuroscientist example

Therefore, we need to take the independent, prior plausibility of hypotheses into consideration, in addition to their likelihood

- This is then the *prior probability* of the hypothesis

Likelihood and prior are the main ingredients in Bayes' rule, which is a theorem of probability theory and thought by many to be a paradigm of rationality.

- This rule tells us to update the probability of a given hypothesis (such as the woodpecker hypothesis), given some evidence (such as hearing some tapping sound) by;
  - considering the product of the likelihood (which was the probability of the evidence given the hypothesis)[2] and the prior probability of the hypothesis (normalized so probabilities sum to 1).
  - The resulting assignment of probability to the hypothesis is known as the *posterior probability*.
  - The best inference is then to the hypothesis with the highest posterior probability.

Of course, the drawback with illustrating the problem as I have done here is that there is no intelligent little person inside the skull consciously performing causal inference. On the story we shall develop, which goes back to Helmholtz, what is really going on is that the neural machinery performs perceptual inference unconsciously. As Helmholtz says about the "psychical activities" leading to perception;

- some quote saying these "psychical activities" are not conscious
- *Unconscious perceptual inference*

While the Bayesian, inferential approach to perception is attractive many questions quickly arise. Straight off, aligning perception with ideally rational, probabilistic, scientific-style reasoning seems rather intellectualist

- Basically, Bayes rule is not intuitive to learn / explicit Bayesian reasoning is hard and if this is the case, then how does the brain perform it if we don't

The contrast to the inferential picture of perception is a picture on which perception, rather than being the upshot of *inferential processes in a hypothesis-testing brain* (top-down), is the result of an *analytic, bottom-up driven process* where signals are

recovered from low-level sensory stimulation and gradually put together in coherent percepts.

- On this alternative, non-inferential approach, perception is driven bottom-up by the features the brain detects in the input it gets from the world.
- Crudely, changes in input drive changes in perception, and so top-down inference in any substantial, normative sense is not needed.
    - top-down vs bottom-up debate

There is much discussion about the relative virtues of the feature detection approach vs. the more inferentialist, Bayesian approach

Perceptual inference and binocular rivalry

On binocular rivalry:

- The neural mechanism behind it is still unknown and it keeps throwing up new and intriguing findings. As Porta delightfully puts it, what makes “visual virtue” alternate between the eyes?
- The brain somehow seems to *decide* that there are two distinct things out there, a face and a house—and perception duly alternates between seeing one or the other every few seconds, sometimes with periods of patchy rivalry in between.

Notice that it (binocular rivalry) puts pressure on the idea that perception is purely stimulus driven, bottom-up feature detection.

- During rivalry, the physical stimulus in the world stays the same and yet perception alternates, so the stimulus itself cannot be what drives perception

To put it in the Bayesian vernacular, the prior probability of such a mishmash cause of my perceptual input is exceedingly low.

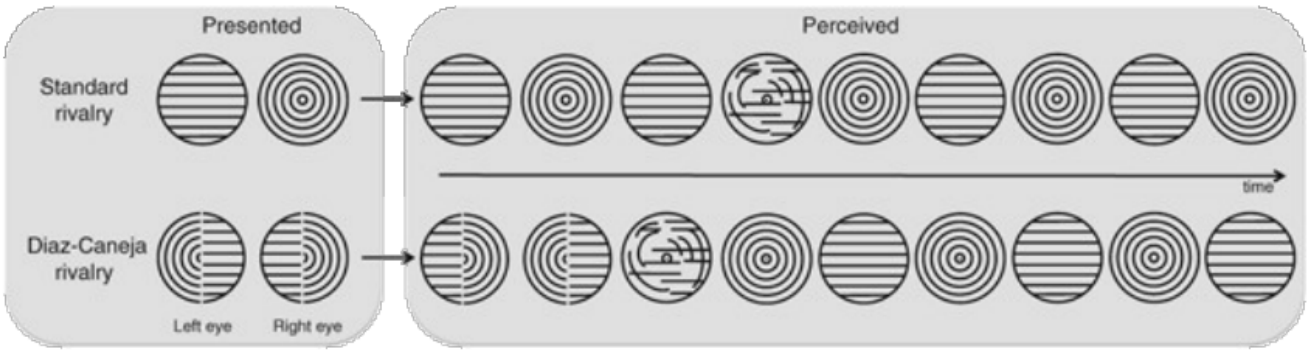
Instead, very “revisionary” hypotheses are selected, each of which effectively suppresses a large part of the incoming sensory signal. It is as if when a face is seen the visual system says “it is most probably a face, never mind all the parts of the total input that the face hypothesis cannot explain”; and vice versa when perception then alternates and the house is seen.

- How exactly this inferential process proceeds is a further matter but it is difficult to see how we could even begin to explain this effect without appealing to some kind of inference.

Recall the worry that the Bayesian, inferential approach to perception seems rather intellectualist. The initial response to this is then that some degree of inference seems to be necessary at least in some circumstances.

- It is of course possible that the brain only has to resort to this kind of inference in special cases like rival input to the eyes.
    - However, it would be odd if the brain had evolved a highly sophisticated inferential process to deal with a perceptual situation it encounters mainly in highly artificial laboratory settings (though there is debate about how uncommon it is, see Arnold 2011; O'Shea 2011)

In 1928 Emilio Diaz-Caneja (Diaz-Caneja 1928) discovered that if the two images are cut in half and combined such that one eye sees, for example, half a house and half a face, and the other eye sees the other halves of the house and the face, then there is not rivalry between what is presented to each eye, there is instead rivalry between the full, uncut images of the face and the house:



- Bottom panel: rivalry with Diaz-Caneja stimuli; the different halves of the stimuli are grouped together, so that perception oen resembles standard rivalry
- It shows that even if rivalry is to some extent is the result of very low-level brute competition between processing from each eye this cannot be the whole story since half an image is taken from each eye and grouped in coherent, rivalling percepts

Howhy/Roepstorff face/house hypothesis

- The Bayesian story then goes like this. The combined face-house blend hypothesis has the highest likelihood, because it accounts for more of the sensory input than the face or the house hypotheses on their own.
- But this high likelihood cannot overcome the exceedingly low probability that a face and a house could co-exist in the same spatiotemporal location (you might on occasion come across a transparent image of a face positioned in front of a house but it is very difficult to conceive of fully opaque faces and houses in the very same location in space).

- So the hypothesis that is selected, and which determines perception, is either the face or the house hypothesis
    - Causing your perception to alternate between the two?!
        - Nope see below

Input: I	
Hypotheses	F+H: "It's a face-house" H: "It's a house" F: "It's a face"
Likelihoods	P(I/F) = P(I/H) < P(I/F+H)
Priors	P(F) > P(H) >> P(F+H)
Perceptual inference	P(F/I) > P(H/I) > P(F+H/I) 

There's apparently some evidence for this Bayesian account

- For example;
    - Zhou Wen and colleagues (Zhou, Jiang et al. 2010) induced binocular rivalry by presenting participants with images of text markers and roses.
    - They increased the probability of it being the roses by adding *olfactory* evidence and letting participants smell roses too.
    - As predicted by the Bayesian story, the participants consequently spent more time perceiving the rose image

More needs to be done on this simple Bayesian account.

- For one thing, it does not explain why there is continued alternation between images in binocular rivalry.
- As presented so far, the account only explains how it can be that only one image is selected for perception. Nonetheless the basic Bayesian idea at least begins to make sense of some essential features of rivalry.

How do neurons know Bayes?

The proposal is that the brain unbeknownst to consciousness is engaged in sophisticated probabilistic reasoning. This may sound as if neuronal populations in the visual cortex and throughout the brain know and apply Bayes' rule.

- Putting things like this carries a risk of what we might call *neuroanthropomorphism*—inappropriately imputing human-like properties to the brain and thereby confusing personal level explanations with sub-personal level explanations.

Very few people would claim that computers do not engage in computation because the hardware inside them does not know the concepts and rules employed in the program.

- Similarly, we should not claim that brains do not engage in probabilistic inference because the neurons making them up do not know Bayes' rule.
- What we should claim, rather, is that we can only understand how computers engage in computation if we understand how the hardware is able to realize the functional roles set out in computer programs.
- Similarly, we can only understand how brains engage in probabilistic inference if we understand how neurons can realize the functional roles set out by forms of Bayes' rule

From inference to phenomenology

To see this it is necessary to appreciate *a hierarchical notion of perception*. Bayesian perceptual inference applies to all levels of sensory attributes, and perception normally simultaneously takes in a wide range of these levels.

- These levels of sensory processing are ordered hierarchically and this is a crucial aspect of the account of the hypothesis-testing brain.

Specifically, this hierarchical notion of perceptual inference seems able to capture something central about perceptual experience, which sets it apart from mere categorization or labelling, namely that perception is always from *a first-person perspective*.

- It is not just that we see a car but that we see it, as a car, from our own perspective.
- Different levels of our perspectival experience change in concert as the movement of eyes, head, or body changes our perspective on the world.
- Perceptual content is embedded in the cortical, perceptual hierarchy and there can be dramatic changes in this content as our first-person perspective changes

A hierarchy of causal regularities

Perception requires us to extract the regularities from the irregularities, the signal from the noise.

- In science this normally happens by controlling for interfering factors in the lab and intervening judiciously in the causal chain.
- In normal perception it mostly happens by keeping track of and modelling relevant interfering factors

Regularities come at different time scales, ranging from tens of milliseconds to hundreds, to seconds, minutes, and upwards towards regularities or rules that are stable over weeks, months, and years.

Fast time-scale regularities include things like how shadows change as you move an object in your hands, slower ones concern the trajectory of a balloon you're trying to catch, slower still concern the way people tend to respond to your requests, and still slower how people tend to vote in years of financial unrest.

Mostly, there is a trade-off between time scale and level of detail. Fast changing regularities are good for detail; slower regularities are more general and abstract.

- This makes sense when we consider what regularities allow us to predict.
- If I want to predict something with great perceptual precision then I cannot do it very far into the future, so I need to rely on a fast changing regularity

Regularities can be ordered hierarchically, from faster to slower. Levels in the hierarchy can be connected such that certain slow regularities, at higher levels, pertain to relevant lower level, faster regularities

- Causal structure and depth is important to perception in at least three ways.

1. Causal interactions are what make perceptual inference difficult by preventing simple one-one relations between the causes of my sensory input and the sensory input itself, discussed earlier in this chapter.
2. Causal interactions between objects, and between the perceiver and objects, shape our first person perspectival experience (for example, the way a shadow may disappear and reveal the true shape of an object when we hold it out in the sunlight).
3. Finally, causal structure allows us to plan our own causal interactions with the world on the basis of what we perceive.

Perceptual variance and invariance

Fast regularities occur in perceptual inference in the shape of the variant aspect of experience: perception captures our immediate and constantly varying first-person perspective.
Every time there is a difference in first-person perspective, for example as your eyes or head move or objects of perception shift around, the brain needs to process fast causal regularities for very basic sensory attributes such as contour, shading, and orientation.

- Some of these changes are suppressed, such as those arising from quick saccadic movement of the eyes.
    - But many changes are consciously experienced, such as those caused by moving your head to scan a scene in front of you.

At the same time, slow regularities occur in perception in the shape of the invariant aspect of experience: perception depends on our ability to abstract from our immediate fluctuating first-person perspective and focus on states of the world that are less sensitive to the concrete way the world is being sampled by the senses right now.

As Edmond Rolls, who is the architect of an impressive computational model of invariant object recognition, puts it concerning visual perception:

> "One of the major problems that is solved by the visual system in the cerebral cortex is the building of a representation of visual information which allows object and face recognition to occur relatively independently of size, contrast, spatial-frequency, position on the retina, angle of view, lighting, etc. These invariant representations of objects, provided by the inferior temporal visual cortex are extremely important for the operation of many other systems in the brain, for if there is an invariant representation, it is possible to learn on a single trial about reward/punishment associations of the object, the place where that object is located, and whether the object has been seen recently, and then to correctly generalize to other views, etc. of the same object. (Rolls 2012: 1)"

The difference between variant and invariant perception, as defined here, is best conceived as a matter of degree, with somewhat vague endpoints.

- As perception becomes more and more dependent on slower regularities it becomes more and more invariant.

- For example, our perception of people as enduring objects is more invariant than our perception of the change of facial features as a person smiles at us.

Message passing between hierarchical levels

A key element of causal structure has to do with the interactions between regularities at different time scales.
This interaction works in a bottom-up fashion such that, for example, the fast changing regularities governing contour, orientation, and so on help you become more confident that what you are looking at is really a nose belonging to an enduring face.

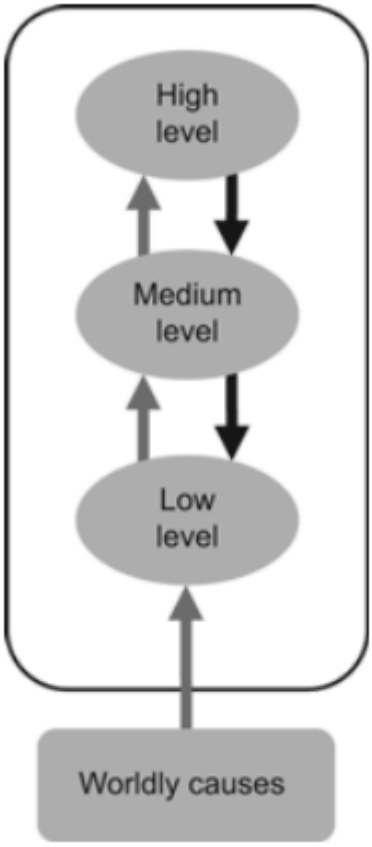
It also works in a top-down fashion such that the longer-term regularities governing faces (e.g., they tend to be attached to headed bodies) assist in recovering fast scale changes in the input from the face (for example, the shadow cast by the nose as the body moves)

Bird song example

- A bird might infer characteristics of another bird through the features of its bird song
    - Forceful/distinct singing might = strength, size e.g.

There is strong top-down modulation of lower level activity.

- That is, variant perception itself is steeped in causal structure. We find it hard to completely divorce the perception of the changing light and shadows from perception of which object it is, and the message passing throughout the perceptual hierarchy reflects this.



- Processing of causal regularities at different time scales influence each other in a bottom-up–top-down fashion.
- Sensory input (dark grey arrows pointing up) is met with prior expectations (black arrows pointing down) and perceptual inference is determined in multiple layers of the hierarchy simultaneously, building up a structured representation of the world.
- This figure simplifies greatly as there are of course not just three well-defined hierarchical levels in the brain.
- A later version of the figure will nuance the description of message passing between levels; in particular, the dark grey arrows will be re-labelled as “prediction error”.

there's more stuff but this is the bulk of it ....

Reading Questions

- What is meant by the phenomenology of perception?
  - the study of subjective experiences as they are directly perceived, mostly concerning the modality of vision
- What is Bayes’ rule, and why might it be relevant to perception?
  - It's a mathematical theorem which is determines how likely a hypothesis is given some evidence. It's relevant to perception because (according to the

article) it's a tool that the brain inherently employs to infer certain information in our perception of the world

- What is the relevance of being able to build invariant representations?
  - They make it possible to learn about reward/punishment associations of an object in a single trial thereby allowing for the brain to correctly generalize this information to other views of the same object, basically greatly lessening the requirement of a perceptual task
- Why would a hierarchy of processing be helpful to perception?
  - It's more efficient to infer some parts of perception (fill in the gaps) because it lessens the amount of processing power required to extract information from the visual input (my interpretation)

Footnotes

1. bootstrapping↩
2. Conditional probability I think ↩

4 Action and expected experience

Predictive processing 1 Perception as causal inference 12 Into the predictive mind

Abstract

The description of the perceptual mechanism has so far been very passive but in fact action is heavily involved in unconscious perceptual inference.

This chapter first connects perceptual inference with action via the notion of active inference.

- An important part of this story concerns the idea that our model of the world includes representation of ourselves.
- Next, however, some issues arise, concerning the notion of self-fulfilling prophecies, and how surprise is bound under a model of the world. which prompt a more involved and challenging information theoretical approach.
- This second part of the chapter serves to show why action is so central to our fundamental understanding of why and how we engage in prediction error

minimization.

This is followed by an exploration of matters arising from this prediction error minimization take on action, including the relation between belief and desire.

The chapter ends on a more general note, by summarizing the prediction error minimization mechanism and commenting on why the framework presented throughout Part I of the book is attractive as well as on what challenges it confronts.

Introduction

A picture has emerged of prediction error minimization as the mechanism the brain uses in perceptual inference. This is in many ways an extremely attractive framework, which includes the following features:

1. It can begin to deal with the problem of perception. The prediction error minimization framework can respond reasonably to the challenge of providing additional constraints on perceptual inference in a non-circular way, without vicious regress, and without being too intellectualist.
2. By appealing to the notion of the perceptual hierarchy, prediction error minimization can begin to accommodate aspects of the phenomenology of perceptual experience, such as the mixture of variant and invariant representation in our first-person perspective.
3. When the notion of precision expectations, and the corresponding notion of gain on prediction error, is built into the mechanism, it becomes possible to see how different overall processing patterns could arise, and how there can be modulation in the engagement of prior beliefs in different situations.
4. The framework is extremely parsimonious, with a simple mechanism at its heart, replicated throughout the hierarchy and yet able to full a number of computational functions.

The key idea in all this is to give the brain, skull-bound as it is, access to not only the incoming sensory data but to a comparison between this data and expectations about what the data should be, under a model of the world.

- The difference between these two is the prediction error, which is then a measurable quantity for the brain, and something that can act as a feedback signal on the way its models of the world are chosen and their parameters revised.

This suggests that the *prediction error minimization* idea should, at the very least, be consistent with the **presence of agency**. But more than that, perceptual inference should be seen as providing an essential part of what makes us beings with **agency**.

- We are engaged in perceptual inference at least in part because we need to act on the world.

In fact, however, there is a much deeper connection between perception and agency, which springs from the very idea of prediction error minimization. Perceiving and acting are but two different ways of doing the same thing.

Active inference in perception

...

The situation is then this.

- Perceptual inference allows the system to minimize prediction error and thus favour one hypothesis.
- On the basis of this hypothesis the system can predict how the sensory input would change, were the hypothesis correct.
  - That is, it can test the veracity of the hypothesis by testing through agency whether the input really changes in the predicted ways.
  - The way to do this is to stop updating the hypothesis for a while, and instead wait for action to make the input to fit the hypothesis.
  - If this fails to happen, then the system must reconsider and eventually adopt a different or revised hypothesis.
    - For example, if the highest posterior goes to the hypothesis that this is a man's face seen in prole, then the system may predict that by moving visual fixation down towards the chin, a sample will be acquired that fits with this hypothesis.
    - If it does, then this further enhances the probability that this is a man's face;
      - if it does not fit then the system may have to go back and revise the hypothesis such that it expects the cause of its input to be, say, a woman's or a child's face

The question arises, why does the system need to engage in this kind of active inference if it has already settled on a hypothesis as having the highest posterior probability? What more is there to do than ranking hypotheses?

- There are two answers to this, which both have to do with reducing uncertainty, that is, with prediction error minimization.

1. Action enhances the posterior confidence in the inference. Action makes decent inferences better.
    - For example, I am more confident I am looking at a man's face after successful active sampling of the world according to this hypothesis. This helps decrease uncertainty especially in cases where the winning hypothesis did not have a very much higher posterior than its competitors at the outset
2. It is efficient and quick to do so
    - Often it is much quicker to form a quick impression of what the ranking might be and then actively test the hypothesis I merely surmise is best.
    - In active testing I can pick a prediction that is made very likely by the hypothesis but very unlikely to occur by chance.
        - If this prediction holds, then the likelihood term is weighted highly and the posterior probability is reinforced.
        - In contrast, in passive observation I have to wait for observations to occur that are strongly predicted by the hypothesis

PLUS

- In many cases of contextual interaction and other causal relations, observation alone will not distinguish between hypotheses where there is causation between two random variables and hypotheses where there is a common cause of covariation amongst the variables.
    - Basically the A/B or C scenario again
- This is a type of uncertainty that can be efficiently reduced by intervening actively
- Given that we can in fact engage in causal inference, and given that observation alone cannot distinguish between such causal models, we can see that we must be relying on intervention, that is, on active inference.

<u>What we have so far is this.</u>

- If the system can act on the world to change its own sensory input, then it can test its own hypotheses.
- It can do this in a Bayesian way by testing primarily those hypotheses that have high posterior probability, endowed from passive perceptual inference.
    - Conditional on the evidence attained in action (for example, as one's eyes move around) a given hypothesis can increase its posterior probability.
    - Through action, already selected hypotheses can be made much more reliable in the sense that they minimize prediction error very efficiently.

- Action in this sense of testing perceptual models is therefore a moment of prediction error minimization—it is active inference

Reading Questions

- If the goal of prediction is to minimize surprise, why act at all?
- Why is it important to have a representation of oneself to perceive the world around one?
- What are the differences between desires and beliefs?
- How do the conception of desires and beliefs relate to the conceptions under behaviourism and computationalism?
- Thinks of some answers yourself to the two challenges to prediction error minimization

12 Into the predictive mind

Predictive processing

Reading Questions

- Is the account of introspection a convincing one?
- What are emotions according to a James-Lange account?
- How is it reconciled that consciousness is private such that it can be social?
- How is the idea of the self as a sensory trajectory related to Kant's idea of transcendental apperception?

Predictive processing

- Perception is causal inference
- Action is active inference
- The mind does prediction error minimization

Learning Goals

1. Understanding how perception can be seen as inference based on sensory input
    - Perceptual inference is done following Bayes Rule

2. Understanding how action is necessary to select the best inference
    - Best inference = the inference that minimises prediction error
3. The capability to reflect on the relevance of this for the symbol grounding problem and artificial intelligence
4. Appreciating how prediction error minimisation may explain mind attributes such as emotion, introspection, privileged access and self

Overview

- Perceptual inference
    - Introduction to Bayes' rule
- Paradigm shift
- Active inference
    - Prediction error minimisation
- Symbol grounding problem
- Emotion, introspection, privileged access and self

Perceptual inference

CLAIM

Perceptual inference is done following Bayes' rule to infer causes of sensory input

MOTIVATION

There is no one-to-one relation between sensory input and what we perceive

Main assumption of predictive processing

> "a [...] substantial view based on the rather uncontroversial idea that the brain is involved in information processing, **and that information theory is cast in terms of the probability theory from which Bayes' rule is derived**" (my emphasis) p. 24

Bayes' rule

- Likelihood
- Posterior probability

- Prior probability
- Marginal probability

Remember

the likelihood principle

O strongly favors H_1 over H_2 iff $P(O|H_1) >> P(O|H_2)$

- $(H|O) \neq (O|H)$
- the likelihood principle is quiescent about $P(H_1)$ and $P(H_2)$

Relation to Methods 2

- Fitting the general linear model

$y = X\beta + \epsilon$
$\hat{\beta} = (X^T X)^{-1} X^T y$ (which is Ordinary Least Squares (OLS))

- the solution that maximizes likelihood

How are likelihood and posterior probability related by? the prior probability
$P(H|O) \propto P(O|H)P(H)$

Hohwy writes: “inference is a *normative* notion” (p. 14). What does he mean by that?

- That it is how it <u>should</u> be, not how it <u>actually</u> is

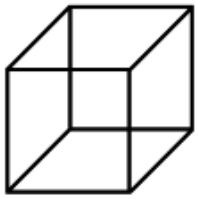
Just mentioning: marginal probability

P(O) is the sum of the probabilities of O conditional on all hypotheses

Y \ X	x_1	x_2	x_3	x_4	$p_Y(y)$ ↓
y_1	$\frac{4}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\mathbf{\frac{8}{32}}$
y_2	$\frac{3}{32}$	$\frac{6}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\mathbf{\frac{15}{32}}$
y_3	$\frac{9}{32}$	0	0	0	$\mathbf{\frac{9}{32}}$
$p_X(x)$ →	$\mathbf{\frac{16}{32}}$	$\mathbf{\frac{8}{32}}$	$\mathbf{\frac{4}{32}}$	$\mathbf{\frac{4}{32}}$	$\mathbf{\frac{32}{32}}$

https://en.wikipedia.org/wiki/Marginal_distribution

Building intuition for “perception is inference”



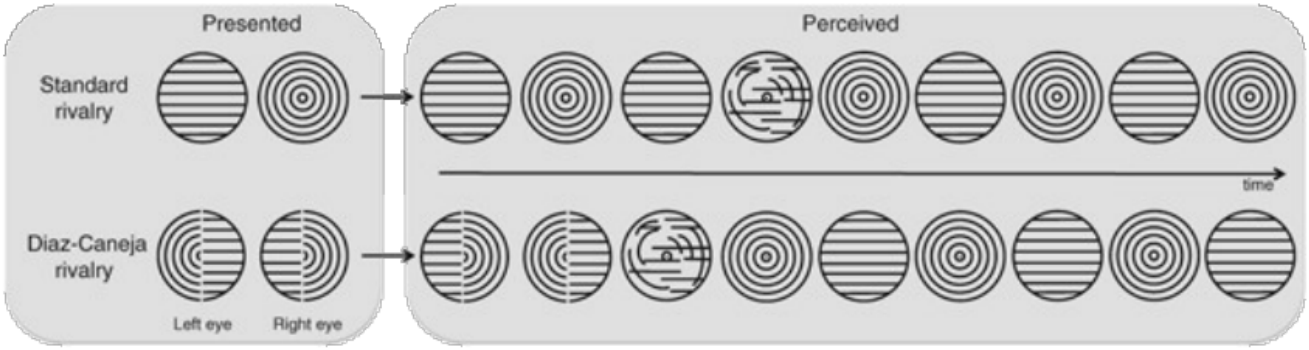
By Cecilia Bleasdale - https://web.archive.org/web/20150227014959/http://swiked.tumblr.com/post/112073818575/guys-please-help-me-is-this-dress-white-and, Fair use, https://en.wikipedia.org/w/index.php?curid=69200610

By Fibonacci - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1788215

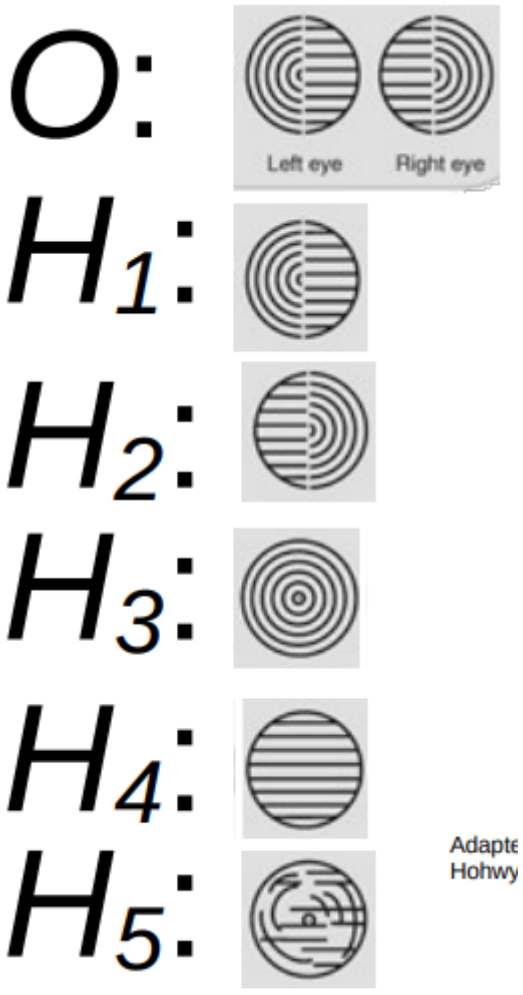
H1: It’s a vase
H2: It’s two faces

- Discuss in pairs: What happens when perception shifts? Explain in Bayesian terms
  - The posterior probability of the respective hypothesis changes

Binocular rivalry



- What do the Díaz-Caneja stimuli reveal about perception that the standard stimuli do not?
  - That there is some prior probability that influence how things behave/what we end up seeing (the brain infers our perception)



- Discuss in pairs: What does the sequence of perception below suggest about how the mind assigns prior probabilities to H1-H5?
  - Our top-down processed ideas of what should appear (prior probability) will decide what will appear to us

Paradigm Shift

> "our conception of the mind is ready to move beyond metaphors of symbol manipulation, connectionism, and dynamics” pp. 24-25"

- dynamics ≈ embodiment
- A very strong claim? Materialistic monism: Identity between brain and mind?

> “the mind [...] is: the result of the dynamics of a complex, physical, information processing system, namely the brain” p. 494

Sounds like T4-indistinguishability – i.e. we need neuromimicry

T4 indistinguishability

Harnad T4

Letting go of metaphors?

- Computationalism ⇒ the mind as a computer
- Connectionism ⇒ the mind as brain
- Dynamicism/embodiment ⇒ the mind as a Watt Governor –
  - just for reference: Van Gelder, T., 1995. What Might Cognition Be, If Not Computation? The Journal of Philosophy 92, 345–381. https://doi.org/10.2307/2941061
- Predictive processing: Explore the mind, by exploring the brain

Remember: Normal Sciences

- A ruling paradigm states the accepted theory of the day
  - Thus also determines the “correct” interpretation of observations

Active Inference

The force of a paradigm

CLAIM

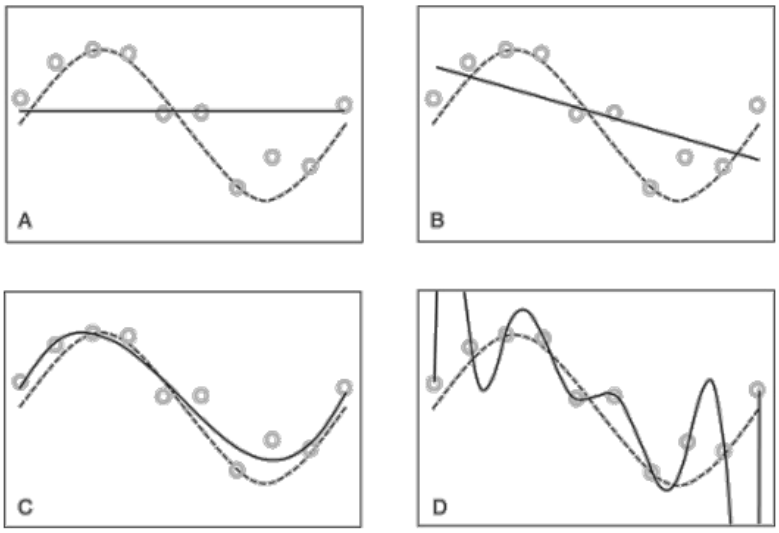
- “The brain is only concerned to minimize prediction error” (my emphasis) p.88

CONSEQUENCE

- All mind/brain functions should/can be understood this way (e.g. chapter 12)

Prediction error minimisation

Discuss in pairs: Which of the four functions is best suited for the data in terms of likelihood, and which is best suited in terms of supplying “tomorrow’s” prior? Bonus question: what are the four functions?



Interim summary

- Bayes’ rule
- Updating your beliefs/hypotheses based on the posterior probability is optimal
- A purely perceptual system faces the problem of how to assign probabilities, P, to hypotheses, H
- Minimisation of prediction error not just a matter of likelihood maximisation, also one of generalisability to new states, O

Active and perceptual inference

- Perceptual inference
  - Models of the world are updated in the light of sensory input
- Active inference
  - We sample sensory input in the light of the model we hold about the world

Remember Kant

> They [Galileo, Torricelli, Stahl] comprehended that reason has insight only into what it itself produces according to its own design; that it must take the lead with principles for its judgments according to constant laws and compel nature to answer its questions [...] -- B XIII

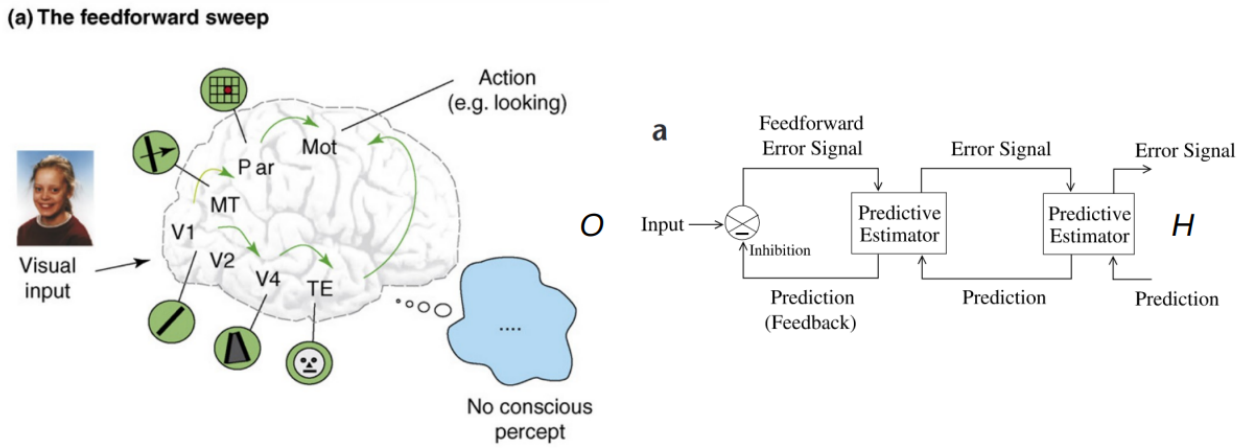
Prediction error minimisation

1. Build model of the world, H, based on sensory input
2. Evaluate H by taking action, thereby getting a new O, using that to update the probability of H
3. Keep taking action to minimise the error on H to build confidence in H
4. If new O indicates that H' is more probable, abandon H, and evaluate H' by taking action

$$P(H|O) \propto P(O|H)P(H)$$

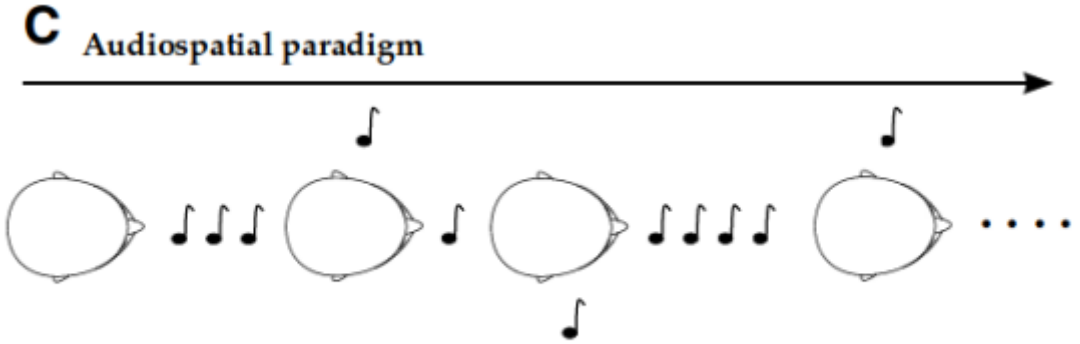
An inversion of the classical conception

VISUAL CORTEX EXAMPLE



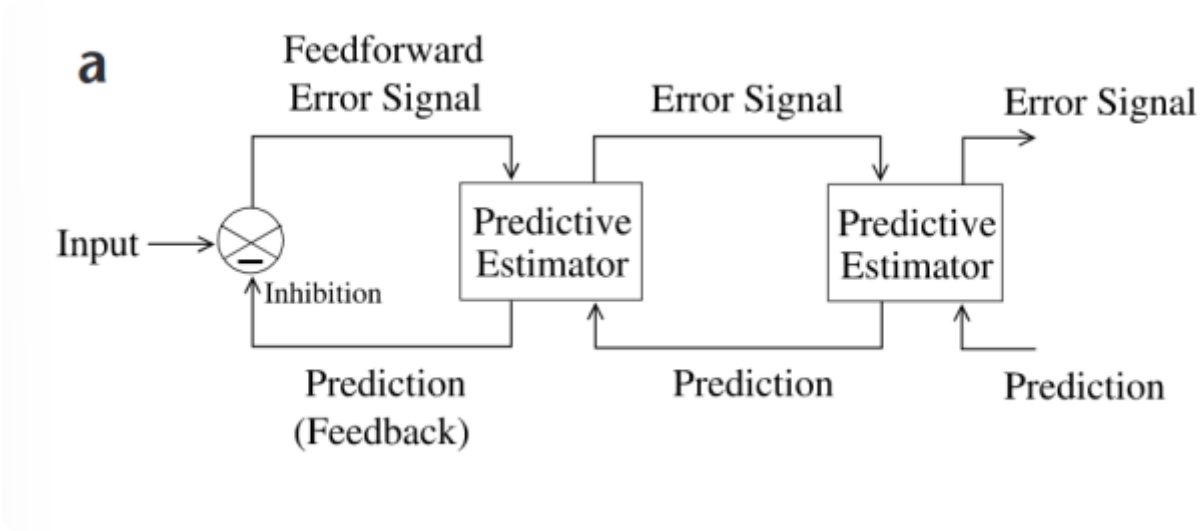
- everything that is already expected is dampened
  - (I don't feel my feet all the time)

auditory example



Symbol Grounding Problem

- In computationalism (and connectionism), the problem was that the meaning had to be “installed” from the outside
  - viz. the Chinese Room
  - and for back-propagation to work in connectionism, having the right answer installed required external interpretations
- Behaviourism on the other hand just did without symbols and representations
  - Skinner argued that these would always require a homunculus inside interpreting them
- We discussed earlier whether sensorimotor capabilities may allow for symbols being grounded
  - Sensorimotor capabilities are necessary for a mind according to predictive processing
    - otherwise, active inference is not possible



- the world itself gives you the truth (it gives the error signals that you need to update your model, i.e. your experience/predictions)

Symbol grounding solution?

> “The world is the truth. [...] Perhaps the mistake, in earlier conceptions of the problem of perception, was to look for a supervisor (a programmer or the system itself), which somehow **knows** the truth [...]. The prediction error minimization approach cuts out the middleman and lets the supervisor be the truth itself.” (original emphasis) p.49

Chapter 12

- Emotion
  - is perceptual inference on our internal states
- Introspection
  - inference on mental causes
- Privacy of self
  - makes my estimate of the state of affairs independent of the estimates of others, meaning that we can use the estimates together socially to make even better estimates
- The self
  - a self-model is needed to predict how acting on the world results in new sensory input

Summary

- Active inference is done to constantly re-evaluate our models/hypotheses of the world
- Sensory input produce prediction errors, which action can help minimise by updating our models/hypotheses
- The symbol grounded problem may be solved by letting the sensory prediction errors be the external interpreter
- Prediction error minimisation is the only thing that the brain does, and it may explain mattes such as emotion, introspection, privacy of self and the self itself

Did you acquire the learning goals?

1. Understanding how perception can be seen as inference based on sensory input
    - Perceptual inference is done following Bayes' rule
2. Understanding how action is necessary to select the best inference
    - Best inference = the inference that minimises prediction error
3. The capability to reflect on the relevance of this for the symbol grounding problem and artificial intelligence
4. Appreciating how prediction error minimisation may explain mind attributes such as emotion, introspection, privileged access and self

Reading q's

● Lamme: – What is meant by phenomenal and access consciousness, respectively? – How would you ascertain the presence of phenomenal experience without report? – Why is the attentional blink classified as a failure to report and not the absence of conscious experience (Table 1)? – Is recurrent processing similar to the hierarchical processing described in predictive processing?

Dehaene: – Is preconscious (Figure 1) similar to Lamme's recurrent processing? – Why is attention necessary to consciousness? (Figure 2) – Why are attention and consciousness different from one another? – What does it mean for content to be globally accessible?

Dehaene - Conscious preconscious, and subliminal processing a testable taxonomy

Consciousness - Lecture & Lamme - Towards a true neural stance on consciousness

Article:
DehaeneChangeuxNaccacheSackurSergent_TaxonomyPreconscious_TICS2006.pdf

Abstract

Of the many brain events evoked by a visual stimulus, which are specifically associated with conscious perception, and which merely reflect non-conscious processing?

Several recent neuroimaging studies have contrasted conscious and non-conscious visual processing, but their results appear inconsistent. Some support a correlation of

conscious perception with early occipital events, others with late parieto-frontal activity.

Here we attempt to make sense of these dissenting results. On the basis of the **global neuronal workspace hypothesis**, we propose a taxonomy that distinguishes between vigilance and access to conscious report, as well as between subliminal, preconscious and conscious processing.

We suggest that these distinctions map onto different neural mechanisms, and that *conscious perception is systematically associated with surges of parieto-frontal activity causing top-down amplification.*

Introduction

Recently, great progress has been achieved by contrasting brain activation images obtained during minimally different experimental conditions, one of which leads to conscious perception while the other does not.

Surprisingly, however no coherent picture has emerged from those experiments. On the contrary, a controversy has arisen, as some studies suggest that consciousness depends mostly on the thalamus and brain stem [1], others on early visual areas [2,3], and yet others on higher prefrontal and parietal association areas [4–9].

An enabling condition: vigilance

The term ‘consciousness’ has multiple meanings, one of them **intransitive** (e.g. ‘the patient regained consciousness’), and the other **transitive** (e.g. ‘consciousness of color’).

- To avoid further confusion, we abandon the term and use ‘states of vigilance’ to refer to the non-transitive meaning, i.e. a continuum of states which encompasses wakefulness, sleep, coma, anesthesia, etc.

Being in an appropriate state of vigilance (e.g. awake rather than asleep) is an obvious enabling condition for conscious processing of sensory stimuli.

- Empirically, awakening into the vigilant state correlates with a progressive increase in regional cerebral blood flow, first in the brainstem and thalamus, then in the cortex with a particularly important increase in prefrontal-cingulate activation and functional connectivity [10].

- Anesthesia, sleep, vegetative state and coma [1,11] are all associated with modulations of the activity of this large-scale thalamocortical network which also shows high baseline activity during vigilant rest [12] and encompasses prefrontal, cingulate and inferior parietal nodes.

...

In summary, vigilance is a graded variable, and a minimum level is essential for placing thalamo-cortical systems into a receptive state.

Early visual activation is not sufficient for conscious report

We now consider the neural bases of the second, transitive meaning of consciousness, which we term ‘access to conscious report’. How do we consciously perceive a visual stimulus?

Many neuroimaging experiments have demonstrated a tight correlation between the conscious visual perception and the activation of striate and extrastriate visual areas [13–18].

Furthermore, extrastriate regions clearly play a causal role in conscious visual perception, because their selective lesioning eliminates the corresponding contents from experience – for instance a lesion of area V4 can destroy color perception in the contralateral hemifield [19].

We argue, however, that early sensory activation is *necessary* but not sufficient for conscious access, because activity in extrastriate visual areas is frequently observed while participants deny having seen any stimulus [14,20– 23].

- When invisibility is caused by masking [20] or by dichoptic stimulation [14] this stimulus-evoked activity remains weak, and one might argue that its small amplitude alone could explain the absence of conscious perception [2,14].
- However, the visual activation evoked by invisible stimuli can also be very strong, for instance when invisibility is caused by neglect [21] or inattention [22,23]

In a recent study of the attentional blink, we observed that up to about 180 ms after stimulus presentation, the occipito-temporal event-related potentials evoked by a invisible word were large and essentially indistinguishable from those evoked by a visible word [23].

- Yet on invisible trials, the participants' visibility ratings did not deviate from the lowest value, used when no word was physically present.
- Thus, intense occipito-temporal activation can be accompanied by a complete lack of conscious report.

Top-down amplification, long-distance reverberation, and reportability

We [4–6] and others [7,8,24] have suggested that, in addition to vigilance and bottom-up activation, a third factor underlying conscious access is the extension of brain activation to higher association cortices interconnected by long-distance connections and forming a reverberating neuronal assembly with distant perceptual areas.

Why would this brain state correspond to conscious access? Neurocomputational simulations show that once stimulus-evoked activation has reached highly interconnected associative areas, two important changes occur:

- (1) The activation can reverberate, thus holding information on-line for a long duration essentially unrelated to the initial stimulus duration;
- (2) Stimulus information can be rapidly propagated to many brain systems

We argue that both properties are characteristic of conscious information processing which in our view is associated with a distinct internal space, buffered from fast fluctuations in sensory inputs, where information can be shared across a broad variety of processes including evaluation, verbal report, planning and long-term memory [25].

Empirically, access of sensory stimuli to conscious report correlates with the activation of higher associative cortices, particularly parietal, prefrontal and anterior cingulate areas.

...

In many of these paradigms, anterior activation is accompanied by an amplification and an increase in functional correlation with posterior stimulus-specific areas [20,26,30].

- Sudden parieto-frontal activation and top-down amplification are two frequent signatures of conscious perception.

Is attention a confound or a necessity for conscious

access?

Some have argued that many of the above neuroimaging paradigms are inappropriately controlled because conscious perception is confounded with increased attention and more extended stimulus processing.

- For instance, a conscious word can be attended, repeated or memorized while a non-conscious word cannot
- Such confounds would suffice to explain the greater parieto-prefrontal activity to unmasked words [20].

For this reason, Tse et al. [18] have argued that one should prefer experimental designs in which attention is drawn away from the stimulus.

- They show that, in such a situation, correlates of stimulus visibility are found solely in occipital areas, *not* in higher associative regions, and therefore argue that the mechanisms of conscious visual perception lie in extrastriate cortex

We obviously agree on one point: it is important to design paradigms in which conscious perception is not confounded with massive changes in overt or covert behaviour. However, this goal has been achieved in several studies.

- In our recent study of the attentional blink [23], for instance, subjects viewed a fixed stimulus and made similar motor gestures on seen and not-seen trials, yet those were still distinguished by strong parietofrontal activation.

We question, however, the proposal that inattention is an appropriate control. Under conditions of diverted attention, such as those studied by Tse et al. [18], even an unmasked stimulus is not guaranteed to be consciously perceived.

- On the contrary, considerable evidence indicates that without attention, conscious perception cannot occur.
  - In the inattentional blindness paradigm, even a 700-ms stimulus presented in the fovea, when unattended, might fail to be seen [33].
  - During the attentional blink, a mildly masked stimulus, normally quite visible, becomes invisible when attention is diverted to another task [23,34].

The relations between stimulus strength, attention, and conscious perception are complex because attention mechanisms can also be activated automatically in a *bottom-up* manner.

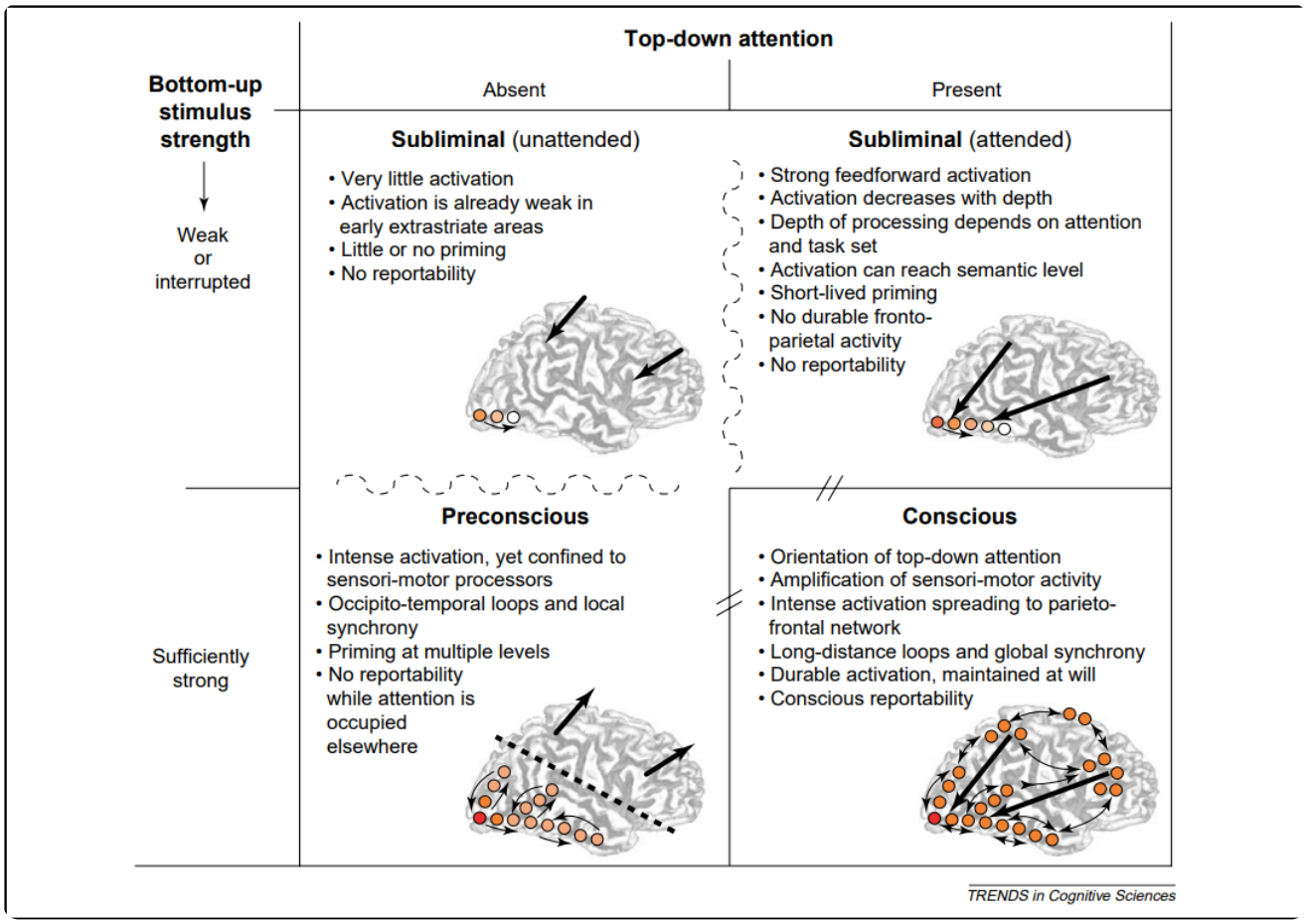
- When the stimuli have strong energy, sharp onsets or strong emotional content, they might trigger an activation of frontal eye fields or amygdala pathways, thus

causing an amplification that can lower their threshold for conscious perception [35].

- Thus, both bottom-up stimulus strength and top-down attentional amplification (whether triggered voluntarily or by automatic attraction) are jointly needed for conscious perception, but they might not always be sufficient for a stimulus to cross the threshold for conscious perception
    - Conscious perception must therefore be evaluated by subjective report, preferably on a trial-by-trial basis.

Distinguishing accessibility from access

The above distinctions lead us to proposal a formal definition of two types of non-conscious processes (Figure 1):



Three types of brain states are schematically shown, jointly defined by bottom-up stimulus strength (on the vertical axis at left) and top-down attention (on the horizontal axis).

- During subliminal processing, activation propagates but remains weak and quickly dissipating (decaying to zero after 1–2 seconds). A continuum of

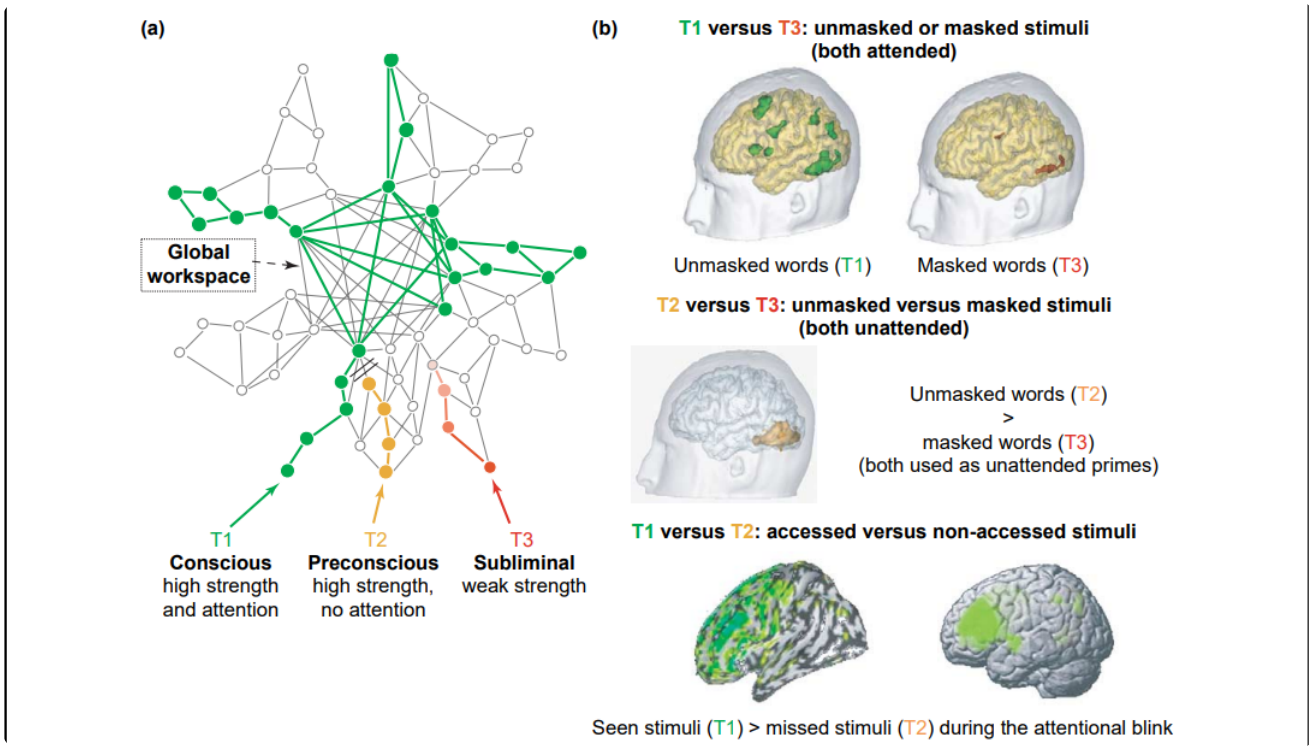
subliminal states can exist, depending on masking strength, top-down attention, and instructions

- During **preconscious processing**, activation can be strong, durable, and can spread to multiple specialized sensori-motor areas (e.g. frontal eye fields).
    - However, when attention is oriented away from the stimulus (large black arrows), activation is blocked from accessing higher parieto-frontal areas and establishing long-distance synchrony
- During **conscious processing**, activation invades a parieto-frontal system, can be maintained ad libidum in working memory, and becomes capable of guiding intentional actions including verbal reports.
    - The transition between *preconscious* and *conscious* is sharp, as expected from the dynamics of a self-amplified non-linear system

1. **Subliminal processing**. We define subliminal processing (etymologically 'below the threshold') as a condition of information inaccessibility where bottom-up activation is insufficient to trigger a large-scale reverberating state in a global network of neurons with long range axons.
    - Simulations of a minimal thalamo-cortical network [4] indicates that such a nonlinear self-amplifying system possesses a well-defined dynamical threshold.
    - A processing stream that exceeds a minimal activation level quickly grows until a full-scale ignition is seen, while a slightly weaker activation quickly dies out. Subliminal processing corresponds to the latter type
2. **Preconscious processing**. Freud [36] noted that 'some processes [.] may cease to be conscious, but can become conscious once more without any trouble', and he proposed that 'everything unconscious that behaves in this way, that can easily exchange the unconscious condition for the conscious one, is therefore better described as "capable of entering consciousness" or as preconscious.'
    - Here we further specify the latter term. We propose to call preconscious (or potentially conscious, or P-conscious) a neural process that potentially carries enough activation for conscious access, but is temporarily buffered in a nonconscious store because of a lack of top-down attentional amplification (for example, owing to transient occupancy of the central workspace system).
    - As shown by the attentional blink and inattentional blindness paradigms, even strong visual stimuli can remain temporarily preconscious.
    - They are *potentially* accessible (they could quickly gain access to conscious report if they were attended), but they are *not consciously accessed at the moment*.

At the neurocomputational level, preconscious processing is proposed to involve resonant loops within medium range connections which maintain the representation of the stimulus temporarily active in a sensory buffer for a few hundred milliseconds.

- A preconscious stimulus might ultimately achieve conscious access once the central workspace is freed (as exemplified by the psychological refractory period paradigm [37,38], in which one task is put on hold while another task is being processed).
- It might never gain access to conscious processing if the preconscious buffer is erased before the orienting of top-down attention (as achieved by masking in the attentional blink paradigm).



Resolving contradictions in neuroimaging studies

global neuronal workspace model A stimulus can fail to become conscious for two reasons

Conclusion

> Instead of the classical binary separation between nonconscious and conscious processing, we introduce here a **tripartite** distinction between *subliminal*, *preconscious*, and *conscious* processing.

The key idea is that, within nonconscious states, it makes a major difference whether stimuli invisibility is achieved by a limitation in bottom-up stimulus strength, or by the temporary withdrawal of top-down attention.

- The first case corresponds to subliminal processing, the second to preconscious processing.

Our proposal could also lead to a reconciliation of several major theories of conscious perception. The distinction between preconscious and conscious processing is consistent with Lamme's proposal of a progressive build-up of recurrent interactions, first locally within the visual system, and second more globally into parieto-frontal regions [3].

Table I. A theoretical taxonomy of conscious and non-conscious information encoding in the brain

Information encoding		Main features
Non-conscious	Latent connectivity patterns	Information is encoded in latent form as matrices of synaptic weights
	Distributed firing patterns	Information is encoded in the distributed firing of many neurons, not condensed in small specialized groups of neurons
	Functionally disconnected systems	Information is encoded in the firing of neurons functionally disconnected from the workspace
	Subliminal processing	Processing is confined to a brief travelling pulse of firing
	Preconscious processing	Processing involves local resonant firing loops, but top-down attention is focussed on another stimulus or task set.
Conscious processing		Processing receives top-down amplification and expands into a global parieto-frontal reverberant state.

Reading Questions

Dehaene:

- Is preconscious (Figure 1) similar to Lamme's recurrent processing?
  - Yes but they disagree on whether or not recurrent processing is the NCC
- Why is attention necessary to consciousness? (Figure 2)
  - Because it is needed to amplify the signal strength of a bottom up stimulus
- Why are attention and consciousness different from one another?
  - Good question
- What does it mean for content to be globally accessible?
  - That it is not limited to local processing in visual regions of the brain but also extended into fronto-parietal areas of the brain thus becoming "global" or globally attended in the brain

Consciousness - Lecture

Run down

- Two important theories of consciousness in cognitive neuroscience will be pitted against one another
  - *Global workspace theory*
    - Conscious items are those we have access to?
  - *Recurrent processing theory*
    - Conscious items are the ones we are phenomenally aware of

Learning goals

1. Understanding the difference between access consciousness and phenomenal consciousness
2. Understanding, how we may able to investigate consciousness in cognitive science
3. Capability to reflect on how consciousness is seen from the viewpoints of behaviourism, computationalism, connectionism and predictive processing

Relation between mind and brain

Overview

- Paradigms
  - are there (conscious) minds?
- How do we (neurally) measure consciousness? ●
- Phenomenal consciousness
  - difference between access and phenomenal consciousness
  - what is it good for?
  - can we measure it?

Are there (conscious) minds

Behaviourism

- Maybe, but only behaviour should be studied and “mind-talk” should thus be translated into behaviour to be studied

Computationalism

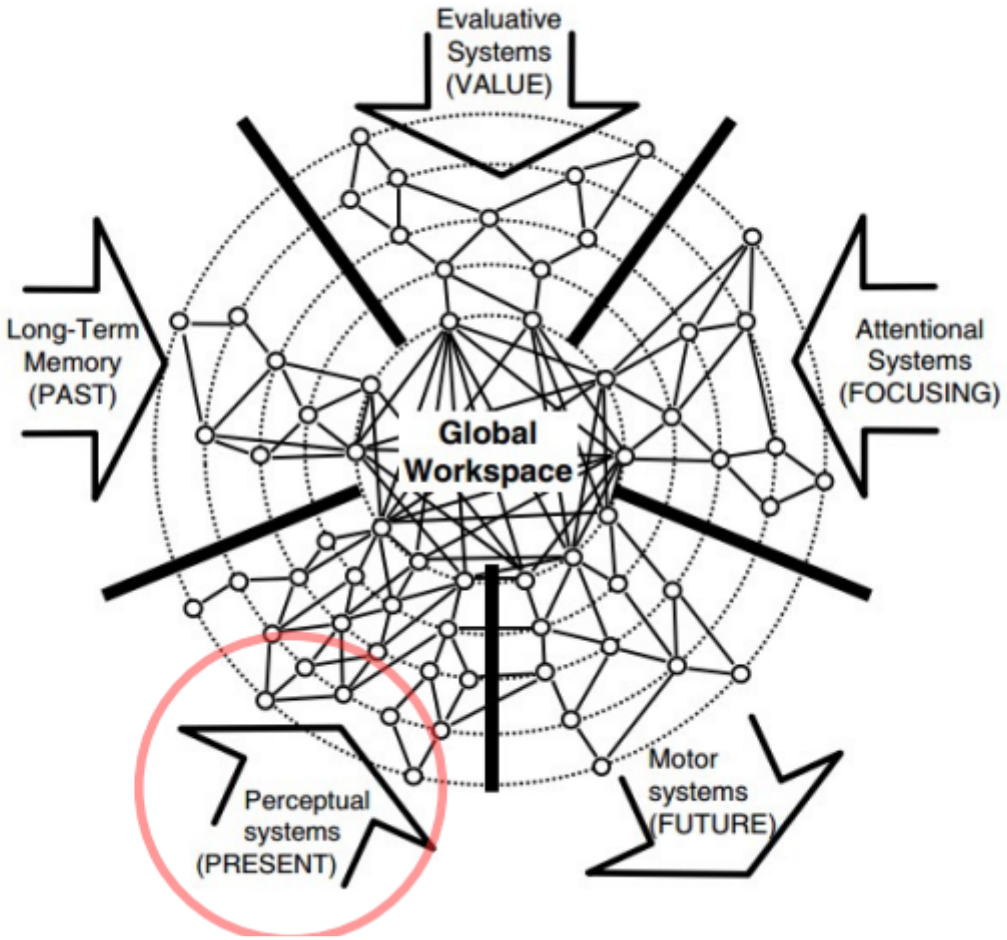
- Yes, and they can be implemented anywhere

Connectionism

- Yes, and for what we know now they depend on brain-like structures

Predictive processing

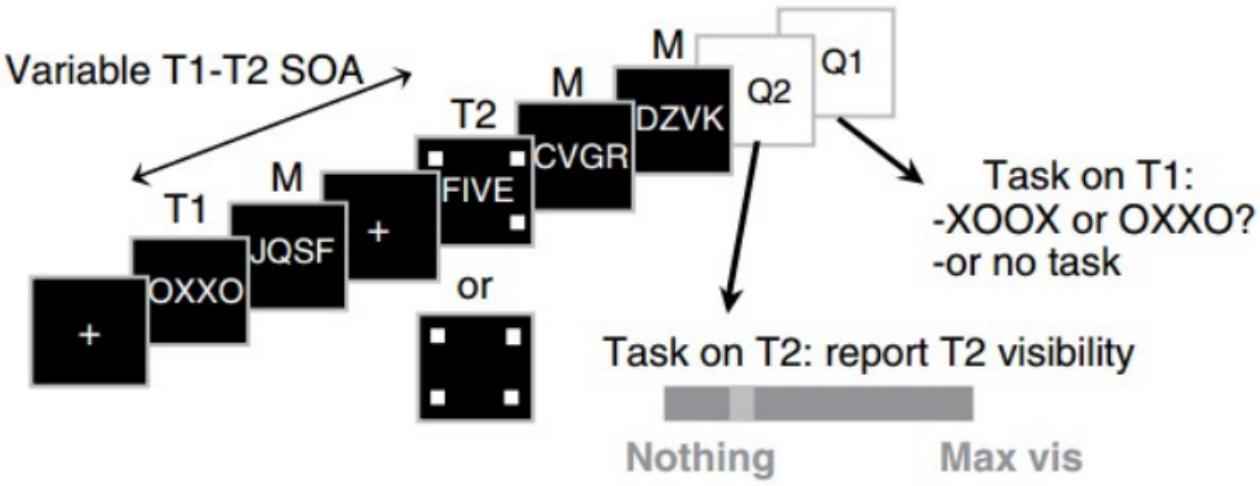
- Yes, and they are dependent on brains that build models of the world (and update them based on prediction errors)



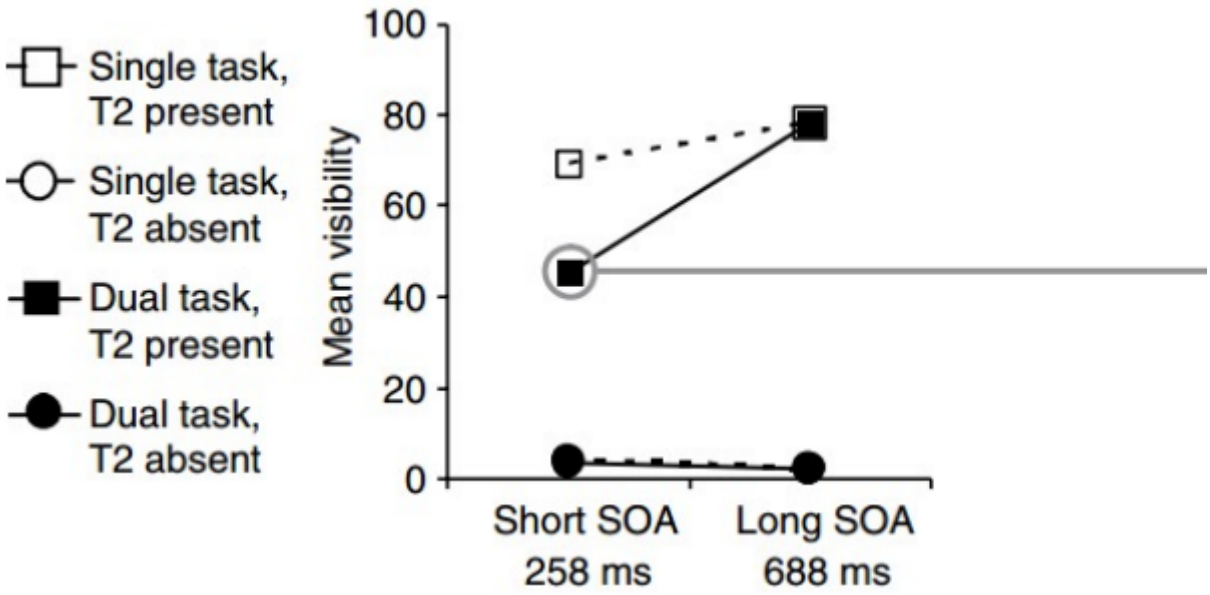
. 2014

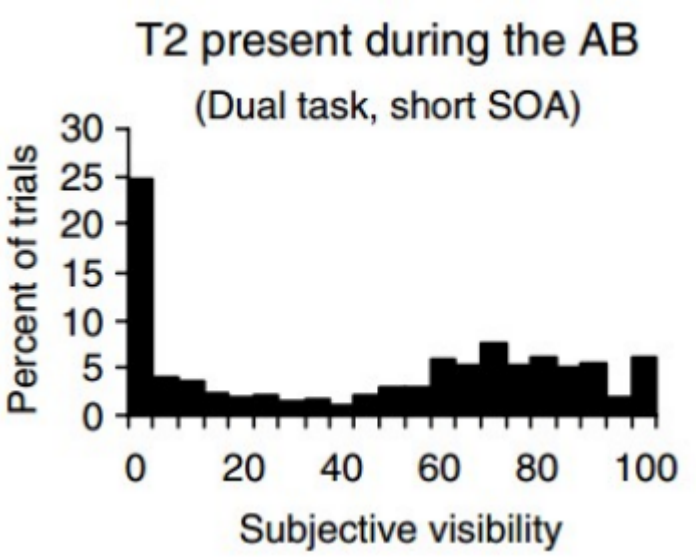
How do we neurally measure consciousness

The attentional blink



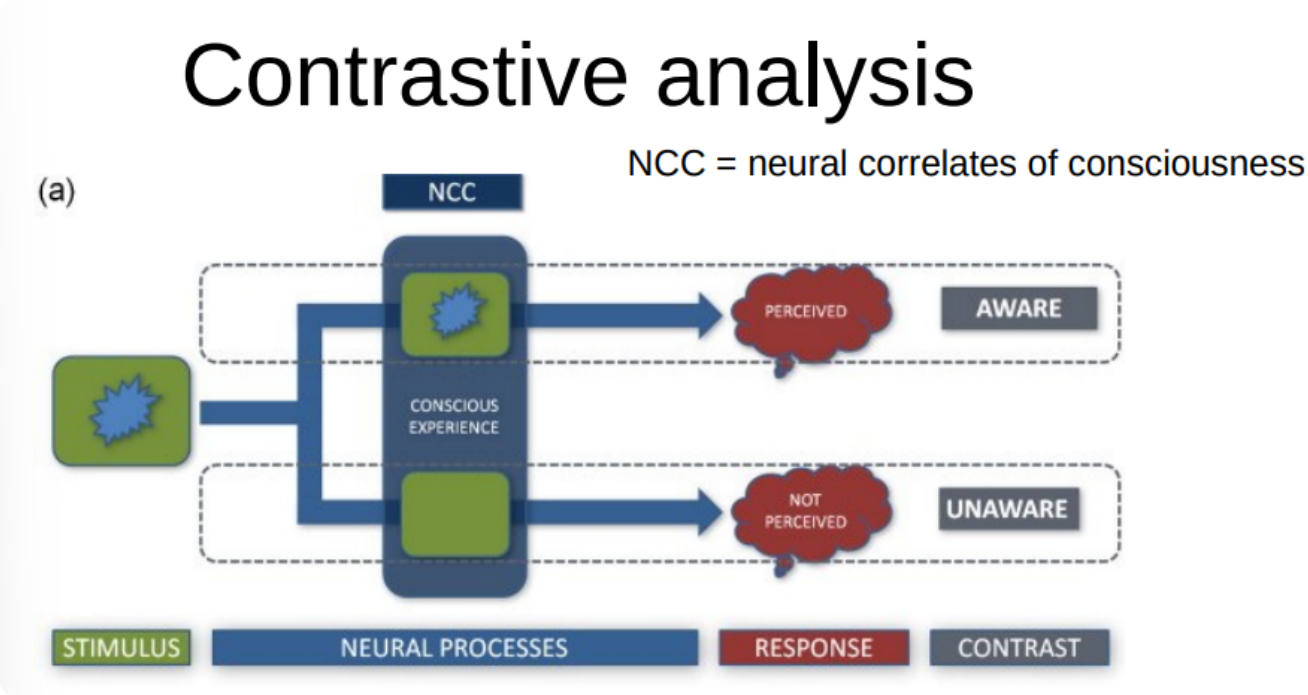
if the duration between these images are too short, people will miss t2
if they don't have a task on t1 then they're more likely to see t2



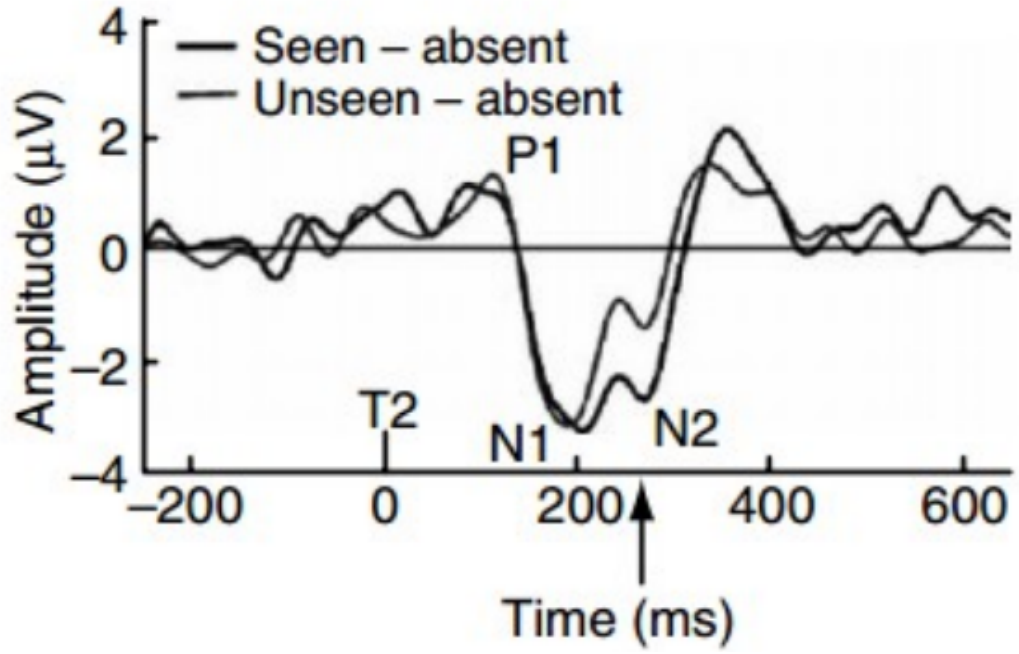


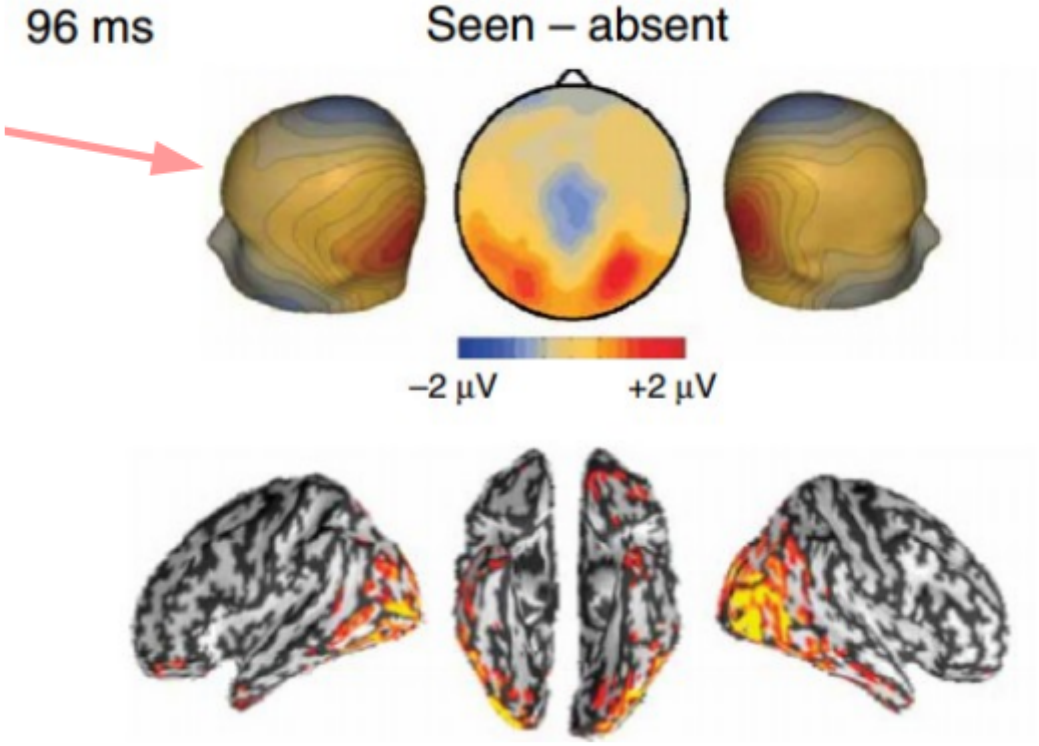
people argue that this graph is bimodal

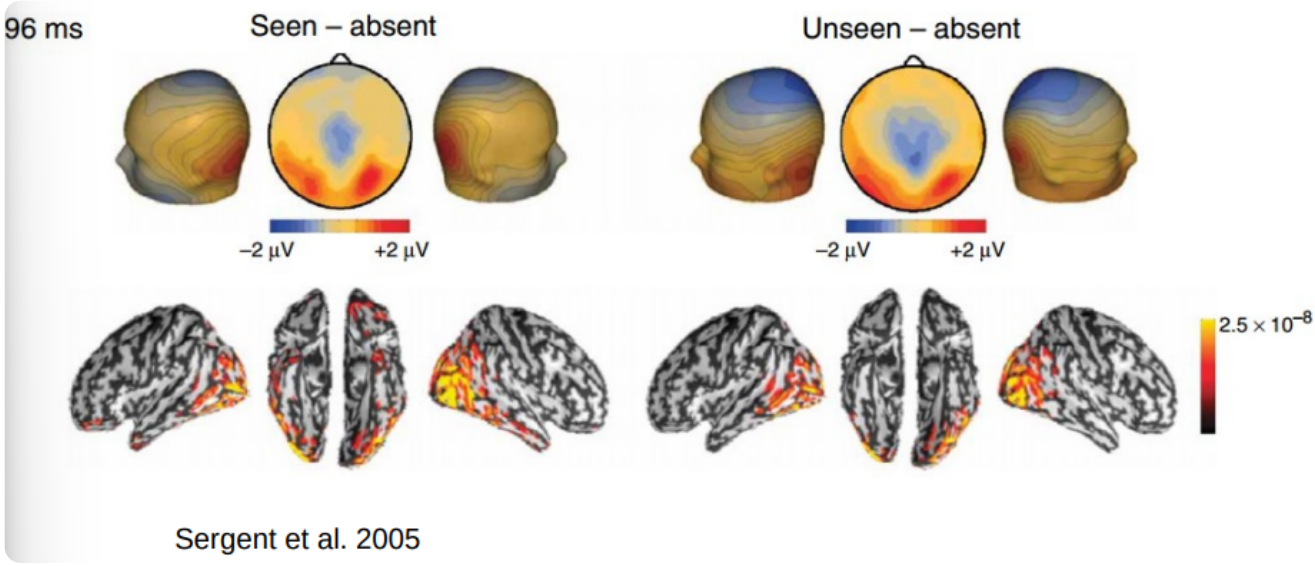
- lau does not think so



- this is how we do neuroscience
- studying contrastts

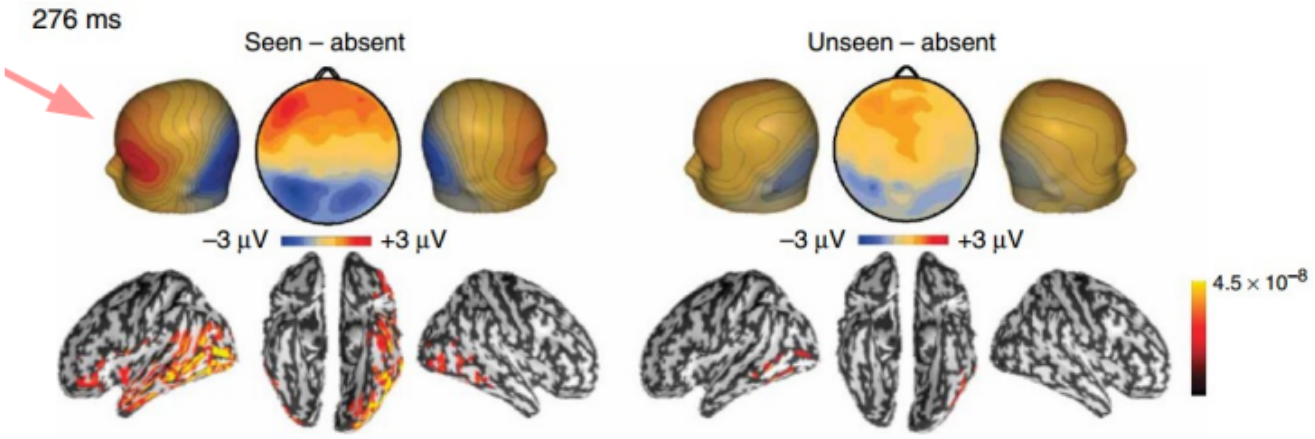

Seen – absent
Unseen – absent
P1
N1
N2
T2
Amplitude (μV)
4
2
0
−2
−4
−200
0
200
400
600
Time (ms)


96 ms
Seen – absent
−2 μV
+2 μV

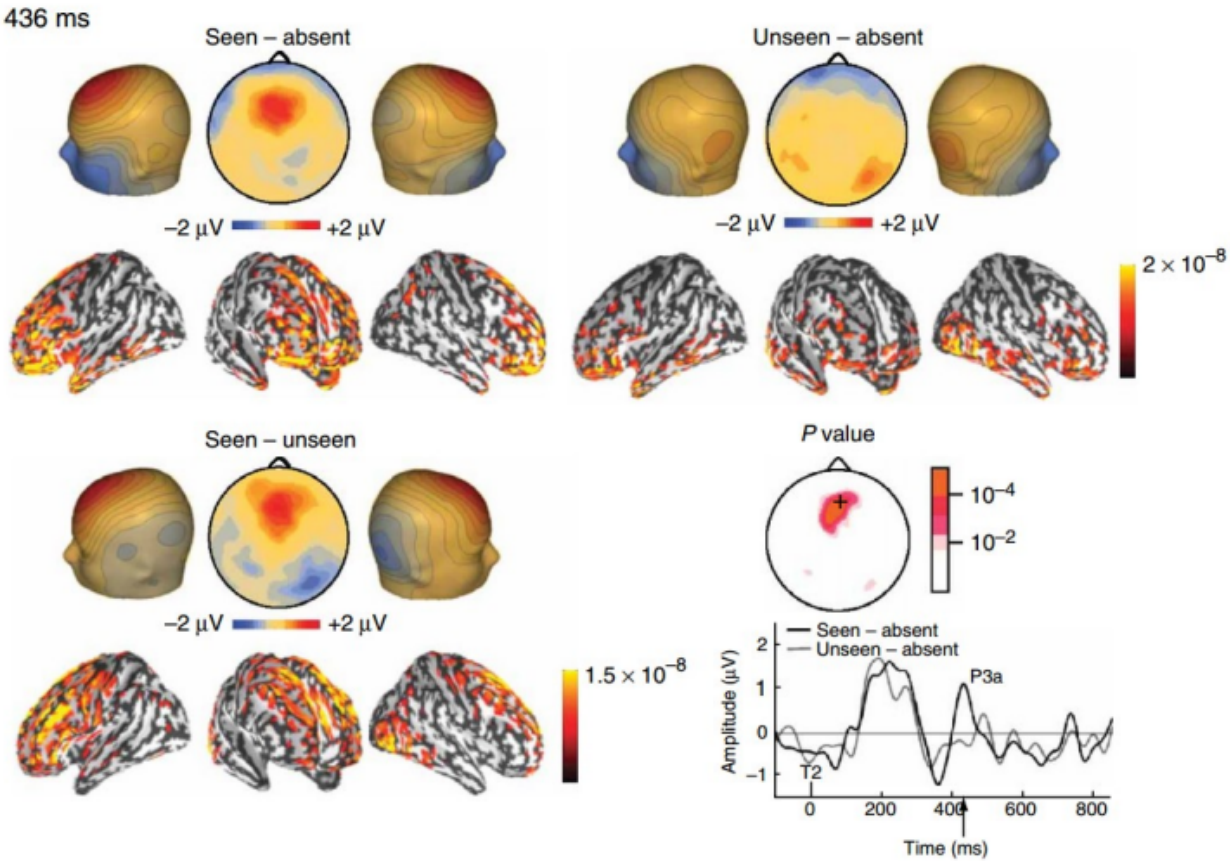


Sergent et al. 2005

- similar early difference waves
- the same visual processing goes on in both of them
  - in one they claim conscious and one they don't
- It means "seen MINUS absent" it's not an "and" sign



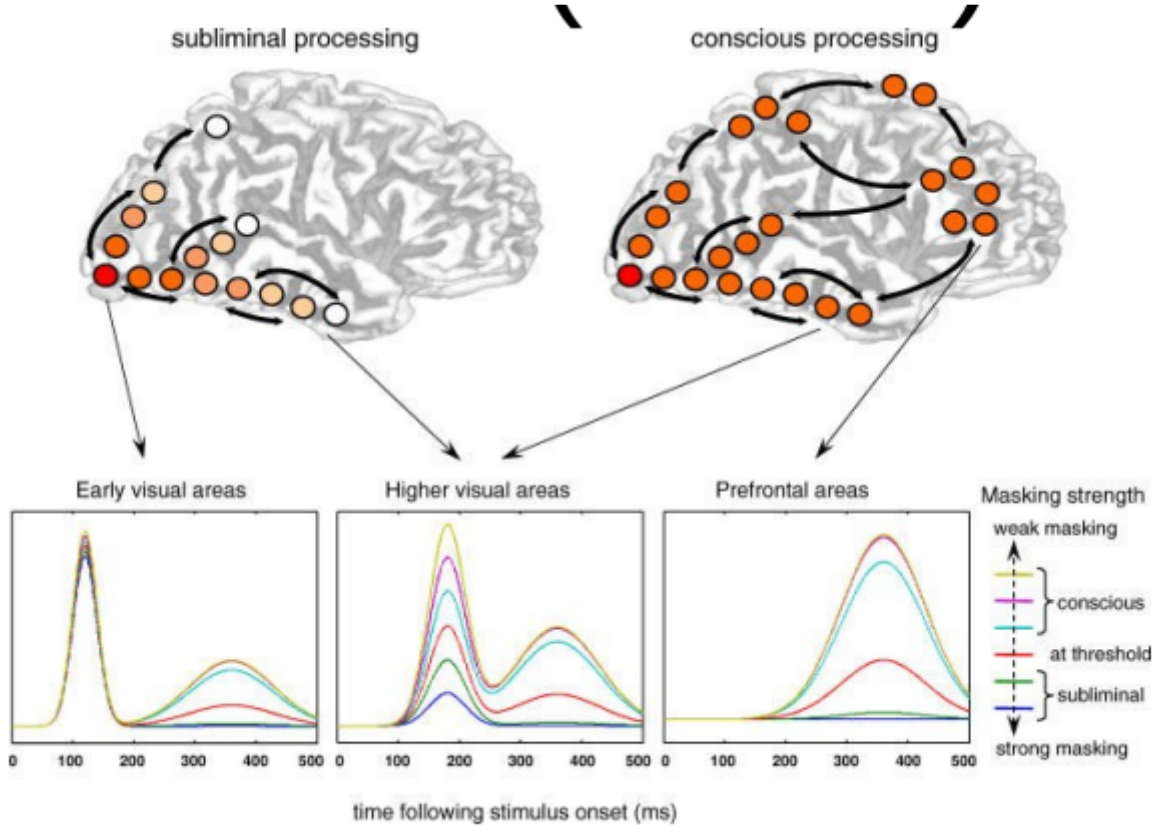
- difference in difference waves, localized to differences in visual cortex
- there's more activity on the unseen now
  - the difference is bigger on the right
  - it seems that we're moving on to something related to consciousness



- difference in difference waves, localized to differences in frontal and visual cortices

one more at 576 ms

> !! Thus, the claim is that the fronto-parietal activation, which there is more of for *seen*'s than for *unseen*'s, activates a *global neuronal workspace* where information can be shared across the whole brain, informing action.



!! So far we have only touched upon the **functions** of the mind, its capabilities of reasoning, thinking, sensing, perceiving etc.

Phenomenal consciousness

what about *experience*?

Thomas Nagel

!! "But fundamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism." (his emphasis, p. 436, Nagel 1974)

!! "We may call this the **subjective** character of experience. It is **not** captured by any of the familiar, recently devised **reductive analyses** of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved like

> people though they **experienced nothing**" (my emphasis, p. 436, Nagel 1974)

Subjectivity is related to point of view and sensory apparatus

- bat example: echo location ("feels" what is there)

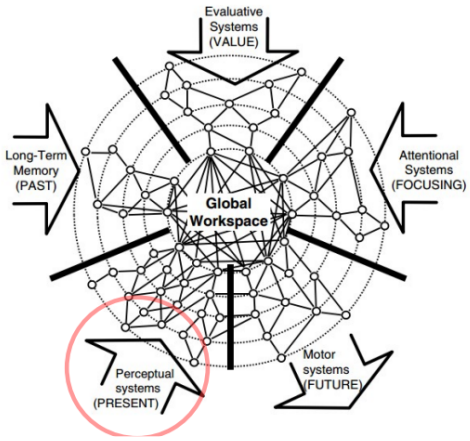
Two kinds of consciousness?

Phenomenal

- related to the content (semantics)
- related to the "feel"
- does not fit an input-output model

Access

- Availability to other systems, e.g. reasoning
- Fits an input output model
- Fits the global workspace model



. 2014

...what we subjectively experience as a conscious state is the global availability of the corresponding information – Dehaene et al. 2014

Interim summary

- We can measure consciousness using contrastive analysis
- For vision, both visual and frontal areas seem to be involved

- Consciousness about content can be divided into access consciousness and phenomenal consciousness
    - Compare with Searle's critique of computationalism

Reprise - are there conscious minds?

Behaviourism

- Translate mental words into behaviour: X desires p, e.g.: If X is in the vicinity of p, X will try to obtain p
- **Issues**: How do we operationalize conscious experience? (I.e. how would it manifest in behaviour?)

Computationalism

- Da Three Levels.png
- **Issues**: Is conscious experience information processing? Does it fit into an input-output model? Why would there be something it is like to process information?

Connectionism

- Build brain-like networks that can emulate mindful behaviour
- **Issue**: Even if we accept that semantics emerge from the distributed network as high-dimensional vectors, how and why would a brain-like model instantiate conscious experience? (symbol grounding problem)

Predictive processing

- Model the mind as a hierarchical structure doing information processing e.g.
    - Interpret perception as providing prediction errors for higher-order models of what the world is like; and allow for action, active inference, to update the models by providing new errors
- **Issue**: Why is subjective experience bound to arise from hierarchical information processing, even if these models are created from a first-person perspective?

The crux of the matter:

- All our paradigms are functional* , i.e. they aim to explain how and why goal-directed behaviour emerges
- However, what is the function of phenomenal consciousness?
    -
        - might predictive processing be an exception?

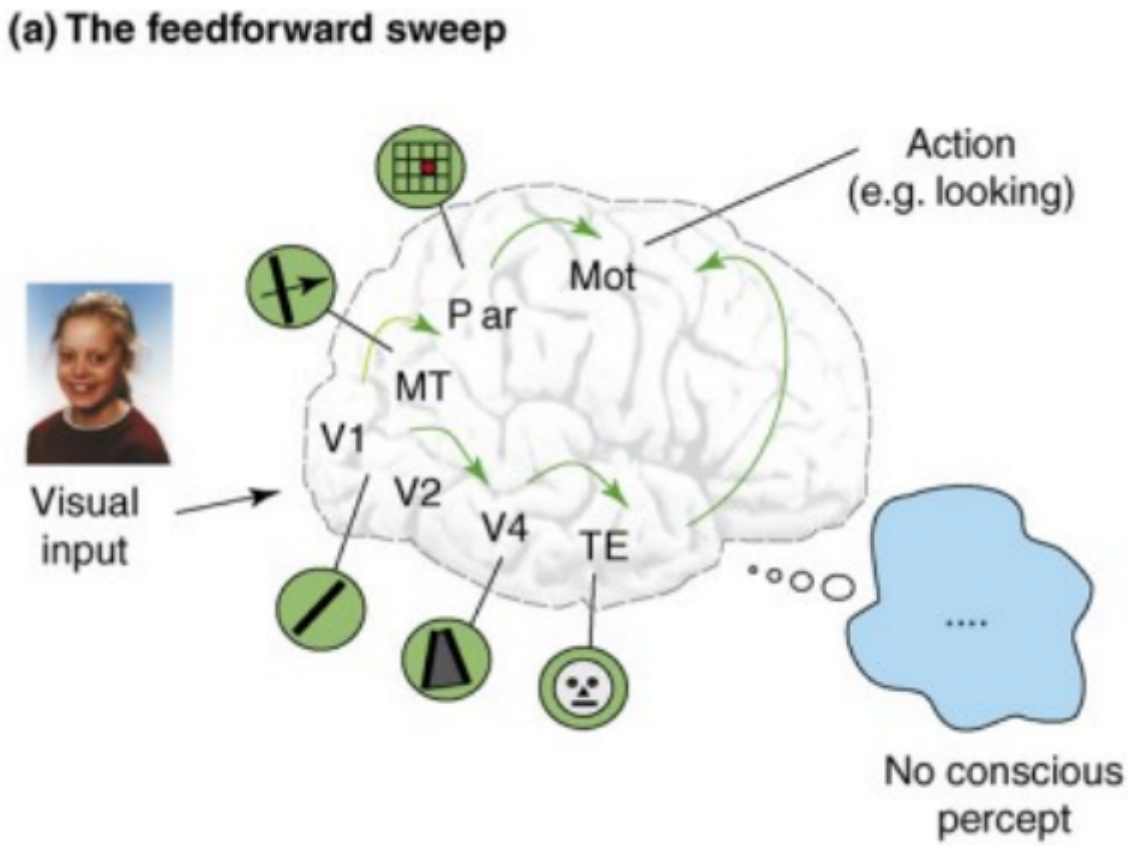
If there is no function...

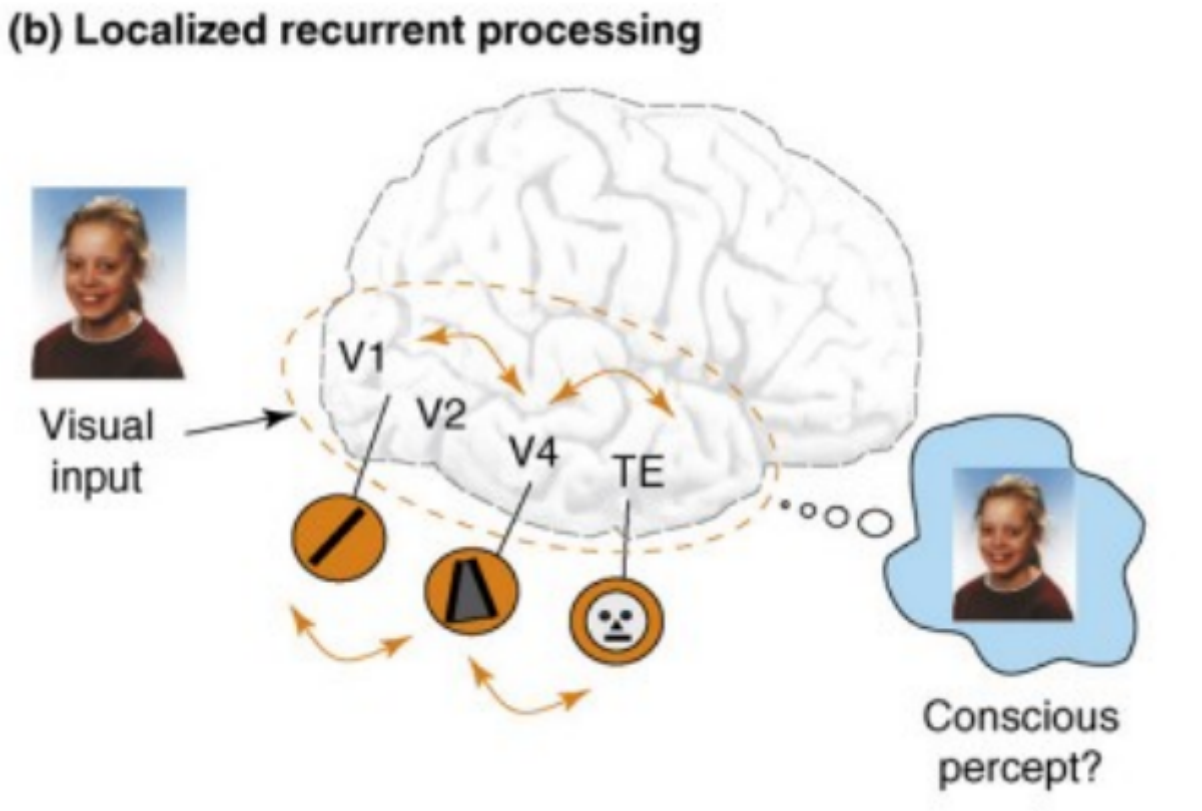
- Consciousness is just an **epiphenomenon**
  - Epiphenomenon: A secondary phenomenon is one that occurs alongside a primary phenomenon (the mind's functional workings), but has no causal connection to world, i.e. conscious experience is a by-product of the mind

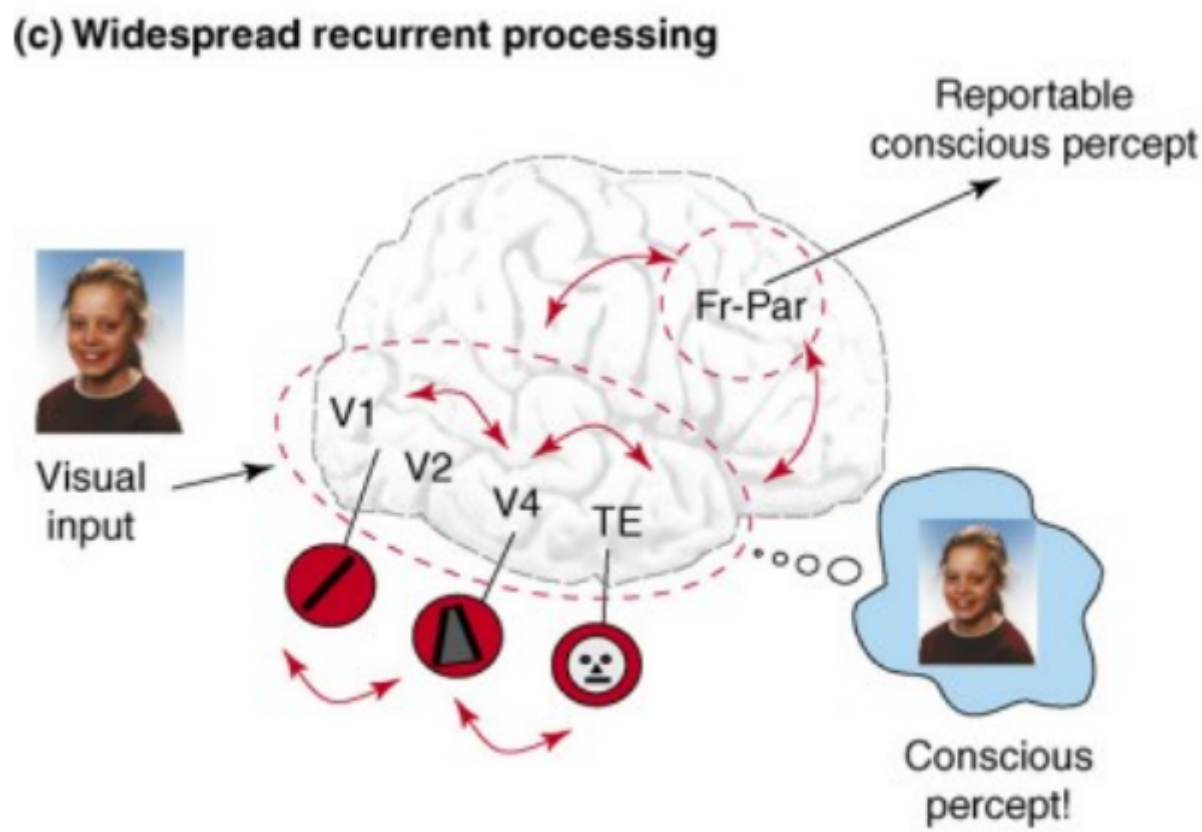
Leaving aside the function, can we, at least, find correlates of phenomenal consciousness using cognitive science?

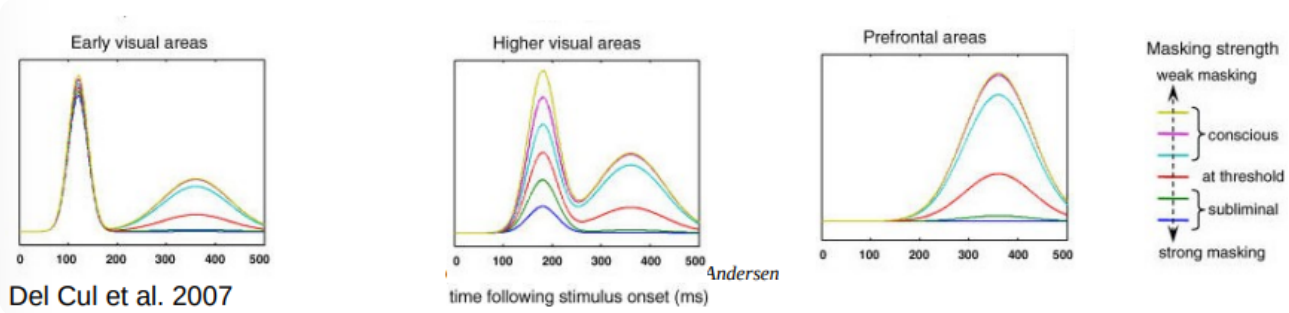
Recurrent processing

- and phenomenal experience




(b) Localized recurrent processing
V1
V2
V4
TE
Visual input
Conscious percept?


(c) Widespread recurrent processing
Reportable conscious percept
Fr-Par
V1
V2
V4
TE
Visual input
Conscious percept!



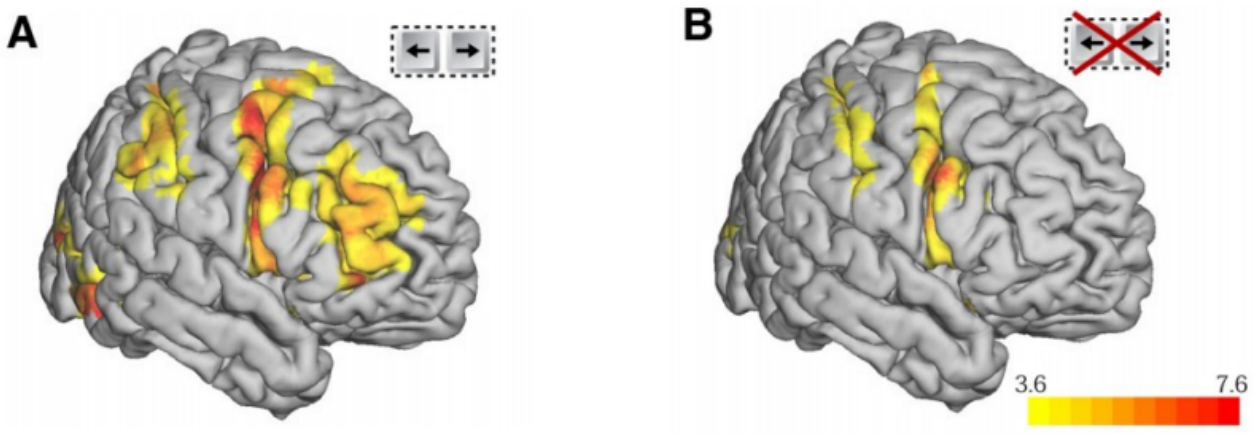
Is report necessary for consciousness?

The attentional blink from Lamme's viewpoint

...

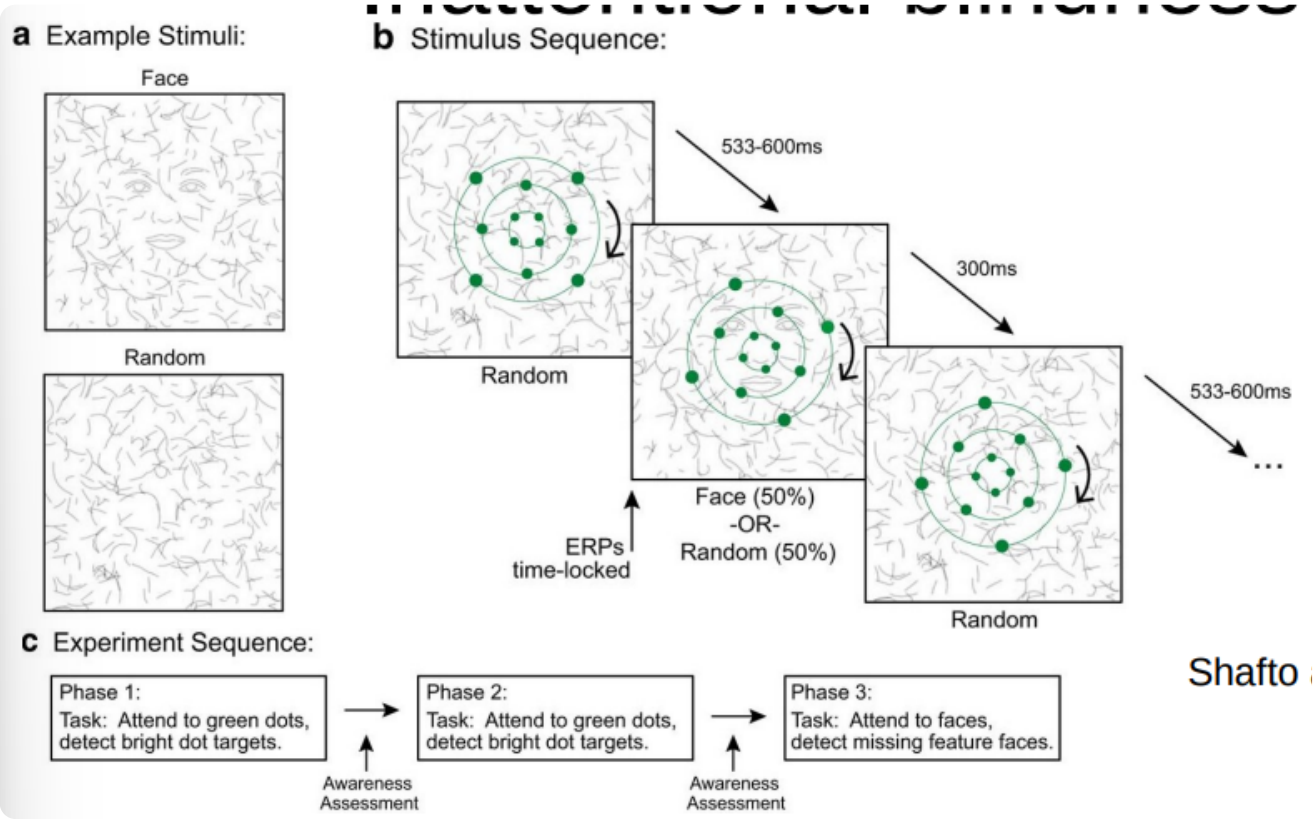
How do we tell if there is a conscious percept, if we don't have our participants report?

binocular rivalry



- frontal regions drop out, without report
  - We have to find ways around having participants explicitly report

Inattentional blindless

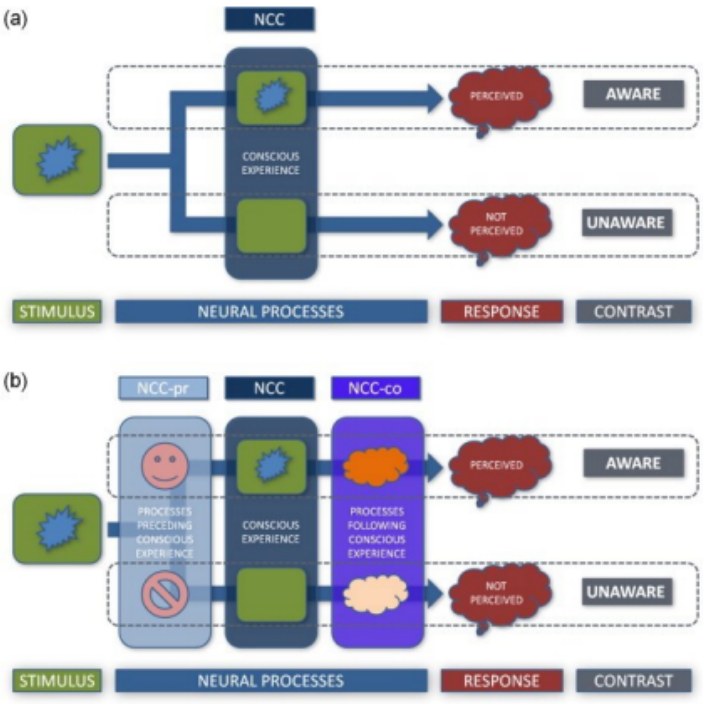


Shafto

...

- is it the act of reporting or actually having consciously seen

too simplistic?



Summary

- We have some ways of evaluating phenomenal consciousness as separate from access consciousness
- It is still an outstanding question what phenomenal experience is good for
- And it is not clear why we would expect something like phenomenal experience arising from how our paradigms model the mind

Disclaimer - what hasn't been covered

- What brings about *state* consciousness
  - e.g., the differences in consciousness when: asleep, awake, in a coma, intoxicated, sleep deprived etc.
  - the self – consciousness about one's self

Learning goals

1. Understanding the difference between access consciousness and phenomenal consciousness
2. Understanding, how we may able to investigate consciousness in cognitive science
3. Capability to reflect on how consciousness is seen from the viewpoints of behaviourism, computationalism, connectionism and predictive processing

Lamme - Towards a true neural stance on consciousness

Same topic:
Consciousness - Lecture & Dehaene - Conscious preconscious, and subliminal processing a testable taxonomy

Article:
towards a true neural stance on consciousness 1.pdf

Abstract

Consciousness is traditionally defined in mental or psychological terms. In trying to find its neural basis, introspective or behavioral observations are considered the gold

standard, to which neural measures should be fitted.

I argue that this poses serious problems for understanding the mind–brain relationship. To solve these problems, neural and behavioral measures should be put on an equal footing.

I illustrate this by an example from visual neuroscience, in which both neural and behavioral arguments converge towards a coherent scientific definition of visual consciousness.

However, to accept this definition, we need to let go of our intuitive or psychological notions of conscious experience and let the neuroscience arguments have their way. Only by moving our notion of mind towards that of brain can progress be made.

Introduction

NCC: Neural correlate of consciousness

Why do some processes in the brain evoke conscious experiences, but others do not?

- neuroscience is expected to provide an answer by finding the neural correlate of consciousness (NCC)
- Experiments to find the NCC invariably involve some manipulation of consciousness, induced experimentally or accidentally, as in lesion patients

Split-brain patients (Table 1) can always report objects presented in the right half field of vision. But when objects are presented to the left, the patients say not to see them, simply because these are processed by the isolated right hemisphere, which has no capacity for language and speech.

- However, many patients can draw these unseen objects, select them from a row of other objects, match them to words, or perform other (simple) cognitive operations, as long as the behavior is executed by the left hand, which is connected to the right hemisphere

Here is the crucial question: Are these patients seeing the objects in the left half field? They say not, and they know best, don’t they? Consequently, we might conclude that the right hemisphere is not part of the NCC for visual experience

- But can you draw something without seeing it?
- And isn’t selection an expression of recognition and attention?

- Isn't it unlikely that the right hemisphere cannot sustain conscious experience, simply because it cannot 'talk'?
- This would argue for *not* excluding the right hemisphere from the NCC.

Some choice has to be made as to what behavioral measures 'count' as evidence for the subject having conscious experience (e.g. drawing versus talk)

In such heterophenomenological observations conscious experience is easily conflated with cognitive functions that are necessary for the report (in this case language).

This problem has grave consequences for the search for the NCC. Many similar examples exist, in which various, if not all, parts of the brain have been included or excluded from the NCC, entirely depending on the measure of conscious experience, or the notion of consciousness that is started with

- in this way, neuroscience will hardly fulfill its promise to get rid of the "tedium of philosophers perpetually disagreeing with each other"

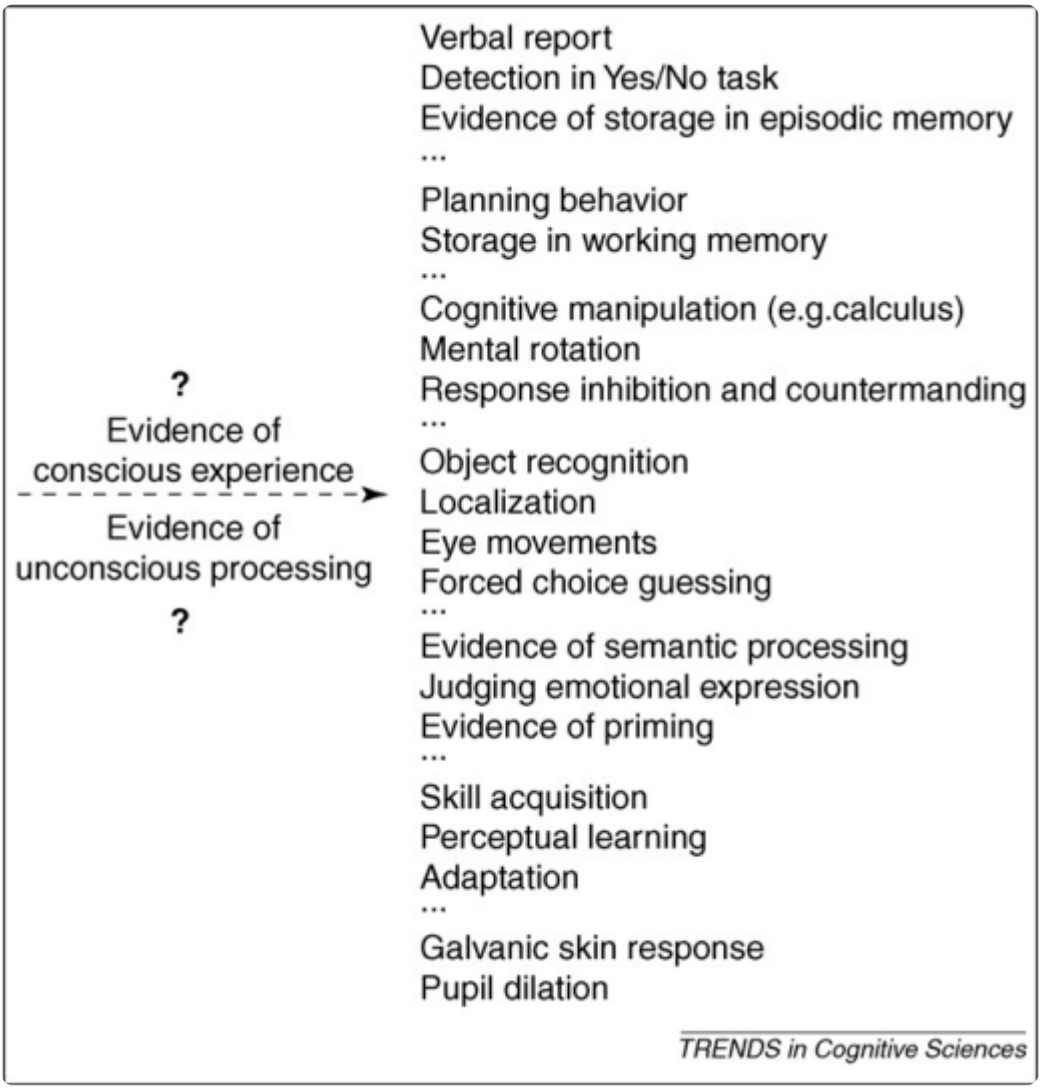
To find a way out, I think we have to go beyond finding 'neural correlates of ', and let the arguments from neuroscience have a true say in the matter.

From neural activation to visual experience

A basic assumption of consciousness research is that when a subject is presented with a visual stimulus it is either seen or not seen. What happens in the brain when a stimulus is shown, and can we establish when conscious experience emerges from the neural activity it causes?

When a new image hits the retina, it is processed through successive levels of visual cortex, by means of feedforward connections, working at an astonishing speed. Each level takes only 10 ms of processing, so that in about 100–150 ms the whole brain 'knows' about the new image before our eyes, and potential motor responses are prepared.

From the very first action potentials that are fired, neurons exhibit complex tuning properties such as selectivity for motion, depth, colour or shape, and even respond selectively to faces. Thus, the *feedforward sweep* enables a rapid extraction of complex and meaningful features from the visual scene, and lays down potential motor responses to act on the incoming information

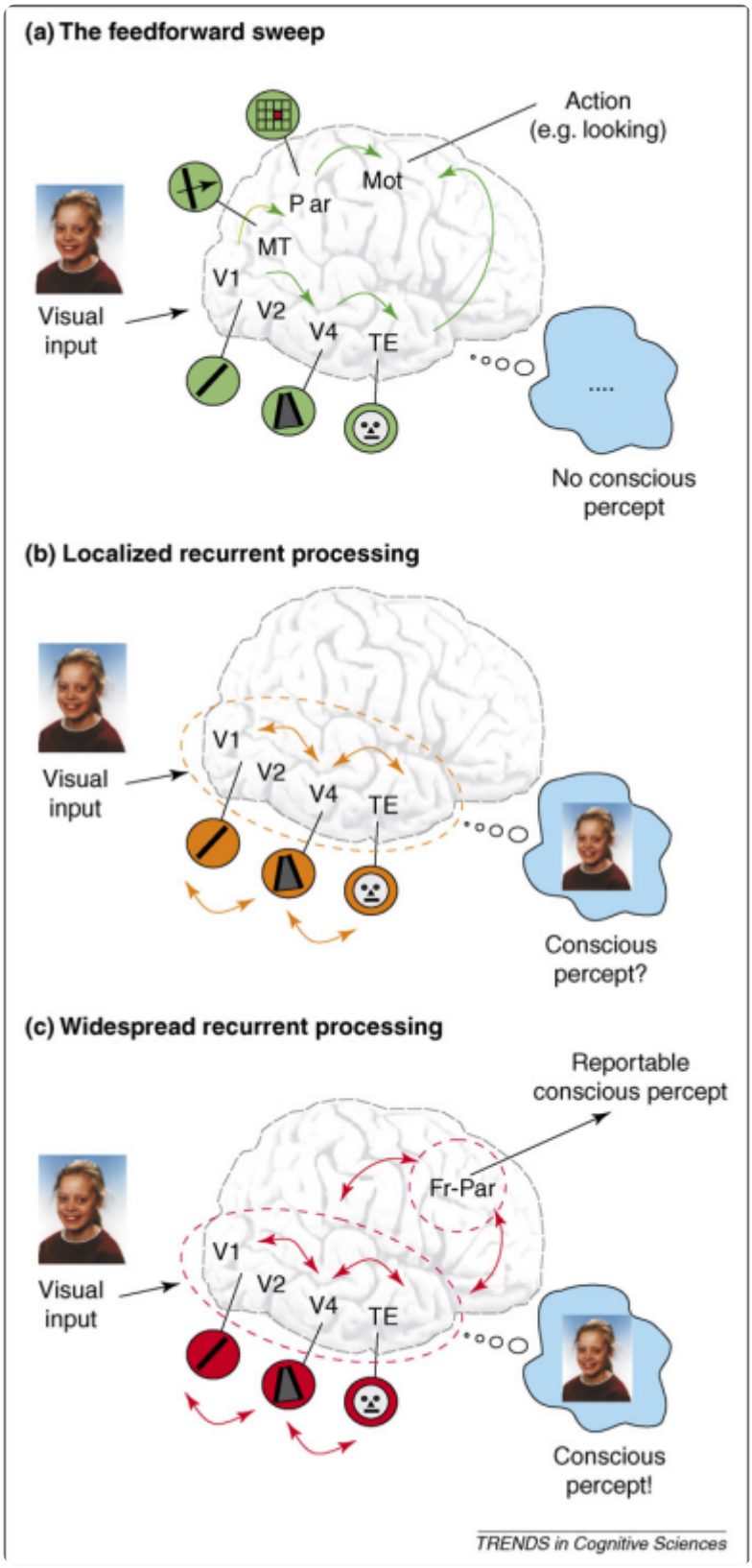


- How to measure conscious experience? How can we tell whether a visual stimulus is consciously perceived? Countless behavioral measures are possible, some of which seem more reliable than others. I compiled a list by asking a group of cognitive psychologists to rank measurements in order of the ‘weight’ these carry as evidence for the subject having a conscious experience.
- Why do some measures end up higher than others? To a large extent, this reflects ideas about the nature of consciousness that have accumulated over years of psychological research and philosophical debate. The problem lies not in the extreme ends of the scale. But where should we draw the line between behavioral measures signaling conscious versus unconscious processing?
- In searching for the NCC, behavioral measures that in one study are taken as evidence of conscious experience are taken as measures of unconscious processing in others.

Are we conscious of the features extracted by the feedforward sweep? Do we see a face when a face-selective neuron becomes active? It seems not. Many studies, in both humans and monkeys, indicate that no matter what area of the brain is reached by the feedforward sweep, this in itself is not producing (reportable) conscious experience.

- What seems necessary for conscious experience is that neurons in visual areas engage in so-called recurrent (or re-entrant or resonant) processing (Figure 2b,c), where high- and low-level areas interact.
- This would enable the widespread exchange of information between areas processing different attributes of the visual scene, and thus support perceptual grouping
- In addition, when recurrent interactions span the entire sensorimotor hierarchy, or involve the frontoparietal areas, potential motor responses could modify the visual responses, which would form the neural equivalent of task set, attention, etc.

That recurrent processing (RP) is necessary for conscious experience is a view that is now being increasingly embraced, although challenging, but entirely empirical, questions remain (Box 1). Box 1.png



- a. The ‘feedforward sweep’, the rapid transfer of visual information through the visual cortex and towards motor areas producing a (potential) response.
  - Within several milliseconds each area extracts information about shape, color, motion, position, objects and faces (see icons). Processing by the feedforward sweep is not, however, accompanied by conscious experience of the visual input.
- b. Recurrent processing enables the exchange of information between higher and lower areas, and within areas, by means of horizontal and feedback connections.

Is recurrent processing between visual areas sufficient for a conscious visual percept?

- c. Reportable conscious experience is present when the visual recurrent core extends towards areas in executive space, such as the *frontoparietal network* (Fr-Par), or language areas

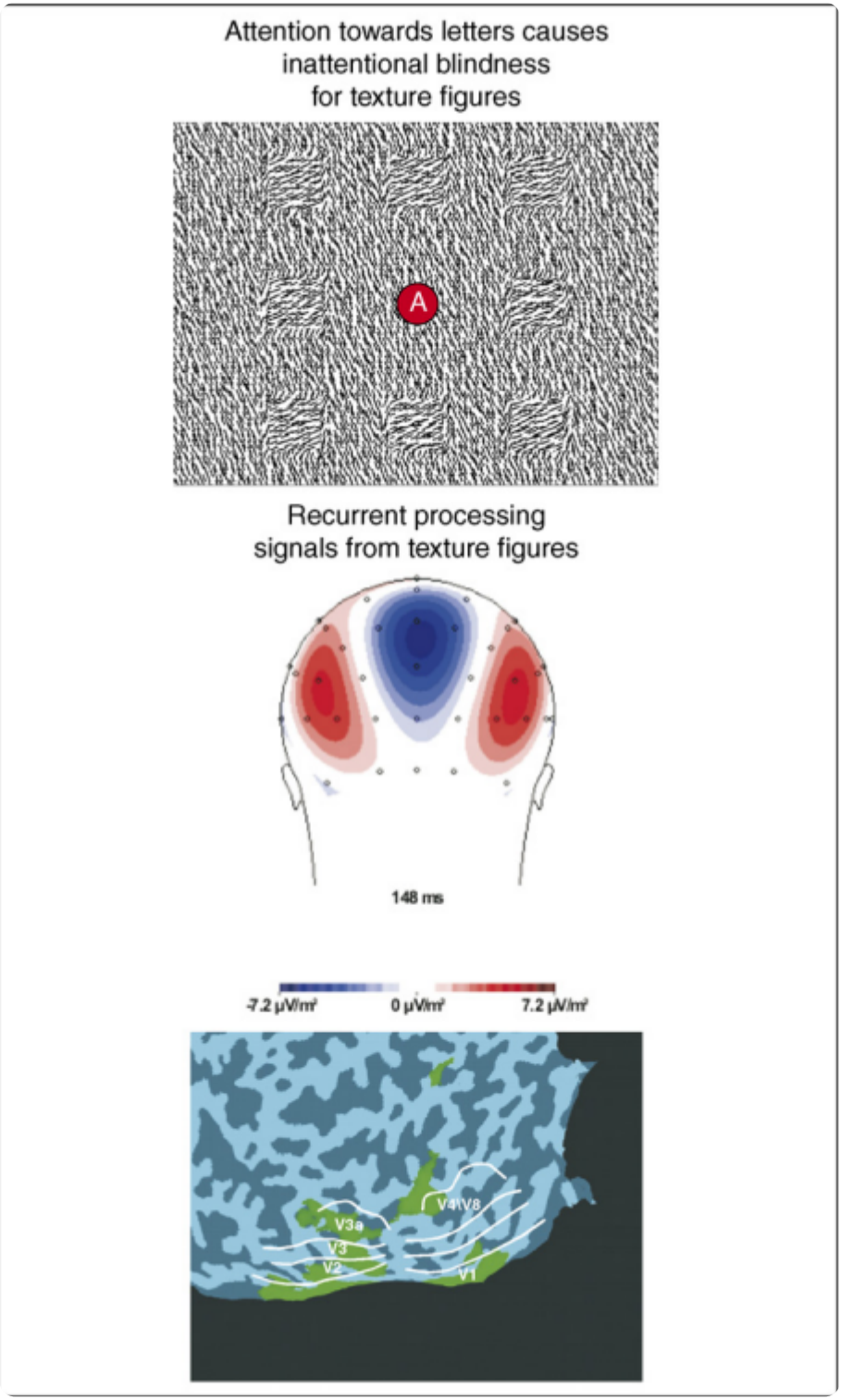
An impossible question?

Much more challenging is another question: is recurrency in itself sufficient for conscious experience?

- Do we see a face as soon as face-selective and lower-level cells engage in recurrent interactions (Figure 2b)?

At first sight, this seems easy to answer. Any instance of RP in the absence of reportable awareness would falsify the idea.

- Super et al. [15] showed that recurrent interactions, recorded in V1 of the awake monkey, are necessary for the animal reporting the presence of texture defined figures.
- However, raising the decision criterion of the monkey, simply by increasing the number of catch trials, *dissociates* RP from the behavioral response
- Then, recurrent signals are also present when the monkey fails to report a figure percept.



- Recurrent processing during inattentional blindness.
- Subjects were instructed to focus attention on a foveally presented stream of black and white letters, and report the occurrence of white vowels.
- All letters were presented in textured backgrounds, but at some occasions, additional square figures were present in the surround of the fixation spot (as shown in the upper panel).
- Afterwards, 50% of subject reported not to have seen these square figures, that is they suffered from inattentional blindness or amnesia.
    - Regardless, brain signals recorded with EEG (middle), MEG and fMRI (lower) revealed that the unseen textured figures evoked recurrent

interactions between several visual areas.

Another study (Figure 3) showed the presence of RP signals, while human subjects were in a state of inattentional blindness (IB) towards the objects inducing these signals. During IB, RP remained localized to visual areas

- Objects were only reported when the signals showed more widespread interactions, possibly involving the frontoparietal network

Does this falsify the thesis that RP is sufficient for conscious experience? Note the similarity with the split-brain conundrum outlined above.

- It could be argued that the only reason why the monkey fails to report its conscious experience is that we have raised the number of catch trials, and hence the monkey's decision criterion.
    - Only if we equate conscious experience to the outcome of his decision process (or to the eye movement signaling that outcome) can we fully 'trust' its report.

Similarly, does IB really show the absence of conscious experience? Simply put, subjects just do not remember stimuli presented outside their focus of attention [17].

- Storage in episodic or working memory is taken as the measure of conscious experience.
    - Unless consciousness and episodic memory are taken as the same thing, it cannot be excluded that the subject was conscious of the stimulus at the moment of its presentation, yet simply forgot about it
- This might sound strange, but experiments on the related phenomenon of change blindness [19] show that unattended objects are represented in the mind in a fleeting, yet reportable, representation that disappears as soon as a new scene is presented or the eyes move to a new location

Again we find ourselves in a labyrinth of arguments

- Take the subjects report at face value and conclude that RP is insufficient for conscious experience
    - In such a view, cognitive functions involved in conscious report (attention, working memory and language) are part and parcel of consciousness, whereas other functions are unconscious
    - *But then why not simply study these cognitive functions, and abandon studying consciousness and the NCC?*
        - Moreover, any sharp division between conscious and unconscious cognitive functions will again be arguable, often resorting to taste rather

than scientific argument.

The alternative view would hold that conscious experience is only done full justice when viewed as independent from other cognitive functions.

- Evidence for this idea is found in the many instances of attention [21], object recognition [22], learning [23], semantic processing [24] and even reasoning [25] without conscious reportability of what was learned, recognized, etc.
- In that view, RP could be seen as the NCC of visual experience, even when not reported.
- But how to verify conscious experience without any cognitive function interfering?
    - This problem lies at the heart of the seemingly unsolvable debate on the existence of 'non-accessible' or phenomenal (P-) consciousness

Recently, the problem was recognized in the context of a global workspace theory of consciousness (Dehaene)

- It was proposed to let such states of localized RP, not involving the frontoparietal networks associated with attention and reportability, fall in a category between the fully unconscious and the fully conscious, called the '**preconscious**', where visual information is 'visible yet not seen'
- This hardly solves the issue of whether there is any phenomenally conscious aspect to such a state, and the authors acknowledge that whether there is conscious experience in such a state 'does not seem to be...a scientifically addressable question

Taking the neuroscience argument seriously

This is where I diverge.

- The question seems not addressable because we do not let all scientific arguments have their way, and treat conscious experience as something that can only be observed behaviorally or introspectively.
- We could find a solution, when we let neuroscience go beyond 'finding neural correlates of '
    - Consciousness should not only be inferred from behavioral measures, to which neural measurements are subordinate. Both measures should be on an equal footing

There is no need to doubt the presence of conscious experience when a subject reports a clearly visible event, nor its absence in cases such as blindsight or deeply masked stimuli (Table 1a).

- In those cases, arguments about the conflation of conscious experience with other cognitive functions do not hold up.
- It is the middle ground where additional, neural arguments become of value (Table 1b).

Let's see what such arguments would bring in answering the seemingly unanswerable: Is there conscious experience when RP is limited to visual areas, as seems to be the case in inattentional or change blindness, attentional blink, and possibly in conditions such as extinction and split brain?

- A starting point would be the observation that when RP encompasses visual as well as frontoparietal areas, a reportable conscious experience ensues [7,10].
- This begs the obvious question as to what the essential neural ingredient of this state is for the *generation of conscious experience*
  - Is it the involvement of the frontoparietal network, as global workspace theory suggests (Dahaene), or is it the recurrency?
  - A neural argument against the former would be that activation of cortical neurons per se, even activation of frontoparietal neurons, is insufficient for conscious experience, as is shown by various masking experiments [14,27–29].
  - This would promote the recurrency as being the most important ingredient.

Another neural argument would come from considering that RP is fundamentally different from feedforward processing.

- RP creates a condition that satisfies the Hebb rule, where the pre- and postsynaptic neurons are active simultaneously
  - This will trigger the activation of synaptic plasticity processes, which are the neural basis of learning and memory [30].
  - In other words, stimuli that evoke RP change your brain, while stimuli that evoke only feedforward activation have no lasting impact.
- It would be difficult to see why the involvement of the frontoparietal network would make such a difference (after all, it is all neurons firing action potentials, whether they are in the back or front of the head). ---- seems like a redundant argument. no?
  - Cases of learning during inattention provide evidence for the point

Theoretical arguments further stress the importance of recurrency. Tononi argues that consciousness is the capacity of a system to integrate information [31,32].

- This capacity depends crucially on the system's propensity to form a 'dynamic core', or complex, characterized by the strong mutual interactions among its elements, as well as by its complexity, that is the large number of possible states it can achieve.
- Because of its connectivity, the *corticothalamic network* has the capacity to form dynamic cores of high complexity, whereas other brain structures, such as the *cerebellum* or *basal ganglia*, do not [32].
    - Therefore, also in this account, recurrent interactions between cortical neurons are the crucial feature of consciousness, not which cortical regions sustain these interactions, which is only relevant for the type of conscious experience that results.
    - Importantly, the proposal allows for multiple complexes to exist at the same time
    - In the IB experiment (Figure 3), one complex could therefore represent the attended stream of letters, while another would represent the not-reported objects in the background. By definition, both would be conscious representations.

According to such empirical and theoretical arguments, RP is the key neural ingredient of consciousness. We could even define consciousness as recurrent processing [10,33].

- This could shed an entirely different light on the matter of whether there is conscious phenomenal experience in states of inattention, split brain or extinction.
- The matter would now become a *scientific debate*, where evidence from behavioral observations is weighed against the evidence from neural measurements.
- If recurrent interactions of sufficient strength are demonstrated, it can be argued that the 'inattentional', 'preconscious' or 'not reported' still have the key neural signatures of what would otherwise be called conscious processing.

What do we lose, what do we gain...

Letting arguments from neuroscience override our intuitive and introspective notion of consciousness seems strange. After all, consciousness is almost synonymous to private access and personal feeling.

So what is the benefit of adopting a more neural stance on consciousness? At first sight, we seem to lose explanatory power. There would be instances of (neurally defined) conscious processing that are not evident from behavioral measures or introspection (Figure 3).

- It is not unusual in science, however, to move away from intuitive notions.
  - This was also the argument in Churchland's response to Searle
- We have grown accustomed to the fact that, despite appearances, dolphins should be called mammals and sodium a metal.
  - There are solid scientific arguments to do so. Are there such arguments in the case of RP and consciousness?

As noted above, RP is crucial for the induction of synaptic plasticity. A first advantage of adopting a neural definition of the conscious–unconscious dichotomy would therefore be that we have a much more fundamental understanding of what consciousness is or does: *we need consciousness to learn.*

- This insight might also open up ways towards an understanding of consciousness at the cellular or molecular level, as synaptic plasticity is associated with specific synapses and receptor types [30,34].
- Although the ‘hard problem’ [1] is not solved by this in itself, it holds better promises of doing so than traditional NCC approaches.

A second advantage is that we would be able to dissociate consciousness from other cognitive functions, such as attention, working memory and reportability, which is a prerequisite for using the term at all

- Elsewhere [10,33], I have shown how, from the neural perspective, attention and consciousness can be orthogonally defined as entirely separate neural processes
  - When other cognitive functions are neurally defined as well, we do not lose explanatory power at all by adopting the neural stance.
  - It can be easily understood why there is no reportability of conscious experience in the case of IB, split brain, neglect and other conditions; in all these cases there are *other* cognitive functions than consciousness (neurally defined) that are manipulated, which is causing the failure of reportability

Third, the neural dichotomy between ‘recurrent = conscious’ and ‘feedforward = unconscious’ yields testable predictions for behavioral experiments.

- If indeed all recurrent processes share the feature of phenomenality and the inclination to induce synaptic plasticity, it can be predicted that learning will follow

the phenomenal aspects of stimuli (e.g. color), rather than their physical features (e.g. wavelength), even when what is learned is not reportable.

Finally, an obvious advantage would be that we can measure the presence or absence of consciousness without resorting to behavioral measures, enabling us to settle long-standing debates about the presence or absence of consciousness in neural disease, locked-in syndrome, coma, anesthesia, or animals.

- In fact, in these cases, neural measures are our only resort.

... by moving our notion of mind towards that of brain?

I have tried to point out that, by adopting a partly neural stance on consciousness, seemingly unsolvable issues in consciousness research can be solved.

- I do not claim that the neural definition of consciousness that I propose [10,33] is ultimately correct. It is the approach that I advocate.

Alternative neural definitions might be proposed [7,35]. Our task should be to evaluate each of these on its scientific merits and capability to deal with all the phenomena concerning consciousness and cognition.

We should be prepared, however, to abandon our traditional ideas about what consciousness is, and let the neuroscience argument have its way.

- Otherwise, it is useless to do any neuroscience on the topic at all.

Reading Questions

Lamme:

- What is meant by phenomenal and access consciousness, respectively?
  - Access consciousness is what we're consciously aware of
  - Phenomenal consciousness is information that is processed in the brain and may or may not become accessible but which we aren't currently consciously aware of
- How would you ascertain the presence of phenomenal experience without report?
  - No idea

- Why is the attentional blink classified as a failure to report and not the absence of conscious experience (Table 1)?
    - The argument is that we're perceptually aware of the stimuli but just don't recruit the required attention towards reporting (because the signal doesn't reach the frontoparietal network)
- Is recurrent processing similar to the hierarchical processing described in predictive processing?
    - Yes very, although

Descartes - Meditations I and II

> !! Cogito ergo sum

Descartes explores the fundamental questions about knowledge, reality and the nature of existence

In the first meditation Descartes begins by expressing doubt about everything he believes to be true

- He introduces the idea of *radical skepticism*, questioning the reliability of his senses and the possibility that an evil demon might be deceiving him

In the second meditation Descartes moves from doubt to certainty

- While he has cast doubt on everything, he realizes that the very act of doubt implies a thinking, doubting self —the famous statement "Cogito, ergo sum" (I think, therefore I am)

Kant, Critique of Pure Reason

> !! In the "Critique of Pure Reason," Kant introduced the concept of *transcendental idealism*, asserting that while we can have knowledge of the phenomenal world (the world as it appears to us), the nature of the noumenal world (the world as it is in itself) is beyond our cognitive reach. He argued that certain fundamental categories and forms of intuition are inherent in human cognition, shaping our experience of the world

Preface to the second edition (B VII- B XIX)

Kant is concerned with the idea that a true science should follow a secure course, characterized by systematic progress, consistency, and shared methodology

In summary, Kant highlights the different paths taken by logic, mathematics, and natural science in their pursuit of knowledge. While logic deals with the formal rules of thinking and progresses securely, mathematics and natural science required revolutions in thinking to establish a secure course as sciences.

Logic as a propaedeutic: "a preliminary study" to the sciences

Reason's Interaction with Nature: Kant emphasizes that reason, in its pursuit of knowledge from nature, must approach it with principles in hand.

- Reason takes the lead with principles for judgments according to constant laws and compels nature to answer its questions.
    - This is in contrast to letting nature guide reason's movements, which leads only to accidental observations without the possibility of connecting them into necessary laws.

Kant is not a fan of metaphysics

- aaAcktualLy!!!
- Kant did not dislike metaphysics, but rather sought to address its challenges and reshape its foundations through a critical examination of the nature and limits of human reason. His work in the "Critique of Pure Reason" had a profound impact on the development of metaphysics and philosophy in the centuries that followed

Kant emphasizes the importance of experimentation in seeking confirmation or refutation of propositions. However, he points out a key difference between natural science and metaphysics.

Challenge in Pure Reason: Kant acknowledges a fundamental challenge in applying experimental methods to pure reason.

- Unlike in natural science, propositions of pure reason, especially those that go beyond the limits of possible experience, cannot be tested through experiments with their objects.
- Pure reason deals with concepts and principles that are a priori, meaning they are known independently of experience.

Feasibility of Experimentation: Kant suggests that experimentation is possible in metaphysics by arranging concepts and principles a priori.

- This involves considering the same objects from two different perspectives: one as objects of the senses and understanding for experience, and the other as objects merely thought for isolated reason striving beyond the bounds of experience.

Two-fold Standpoint: By adopting this two-fold standpoint, Kant proposes that we can experiment with concepts and principles of pure reason.

- ==When we find agreement with the principles of pure reason from both perspectives, it strengthens the credibility of those principles. ==
  - However, if an unavoidable conflict arises when considering the same objects from only one standpoint, the experiment becomes a means to decide the correctness of that distinction.

Copernican Revolution

Kant famously compared his philosophical revolution to the Copernican Revolution in astronomy.

- Copernicus revolutionized astronomy by proposing a heliocentric model of the solar system, with the sun at the center and the planets, including Earth, orbiting around it.
- Similarly, Kant aimed to bring about a revolution in metaphysics by shifting the focus from an objective reality existing independently of human cognition to the idea that human cognition actively contributes to shaping our experience of reality

The traditional view in metaphysics, which Kant termed "*dogmatic metaphysics*," assumed that the human mind passively receives and reflects an objective reality as it is.

- Kant challenged this view, asserting that the mind actively organizes and structures sensory experiences, contributing essential elements to our knowledge of the world
  - Rather than seeing the mind as a passive receiver of information, he emphasized its constructive role in organizing sensory data
  - Kant introduced the concept of a priori knowledge, suggesting that certain knowledge is not derived from experience but is inherent in the structure of the mind

Transcendental Idealism: Kant's philosophy, often labeled as transcendental idealism, posits that while we can know the world as it appears to us (phenomena), the nature of the world in itself (noumena) is beyond our cognitive reach

Introduction (as in the second edition): Sections I-V

I. On the difference between pure and empirical cognition

- " Among a priori cognitions, however, those are called pure with which nothing empirical is intermixed . Thus, e.g., the proposition "Every alteration has its cause" is an a priori proposition, only not pure, since alteration is a concept that can be drawn only from experience."

II. We are in possession of certain a priori cognitions, and even the common understanding is never without them.

- Necessity and universality are secure indications of an *a priori* cognition and also belong together inseparably
- Kant asserts that there are fundamental principles, such as those found in mathematics or the idea of causation (much opposed to Hume's proposal, which he calls "a merely subjective necessity") which are inherent in human cognition

III. Philosophy needs a science that determines the possibility, the principles/ and the domain of all cognitions a priori.

The unavoidable problems of pure reason, according to Kant, are God, freedom, and immortality.

- These issues lie beyond the scope of empirical experience, and reason engages in speculative investigations to address them.
- The science directed toward solving these problems is called metaphysics.
- Kant notes that metaphysics often begins in a dogmatic manner, confidently taking on the task without prior examination of reason's capacity for such an undertaking
- Kant uses mathematical cognition as an example of successful a priori knowledge
- Kant criticizes the common tendency of human reason to construct elaborate philosophical edifices without thoroughly examining the foundation on which they are created

IV. On the difference between analytic and synthetic judgments.

Analytic Judgments:

- In analytic judgments, the connection between the subject (A) and the predicate (B) is thought through identity.
- The predicate B is (covertly) contained in the concept of the subject A.
- Analytic judgments are those in which the predicate is already part of the subject's concept, and the judgment serves to clarify or analyze the concept.
  Example: "All bodies are extended."
- This is an analytic judgment because the predicate "extended" is already contained within the concept of a body. No information beyond the concept is needed to understand this judgment.

Synthetic Judgments:

- In synthetic judgments, the connection between the subject (A) and the predicate (B) is thought without identity.
- The predicate B lies entirely outside the concept A, even though it stands in connection with it.
- Synthetic judgments add something to the concept of the subject, introducing a predicate not already contained in the subject's concept.

  Example: "All bodies are heavy."
- This is a synthetic judgment because the predicate "heavy" is not contained in the concept of a body. It adds new information to the concept through a connection established by experience.

There is more but it honestly seems somewhat unnecessary
https://chat.openai.com/share/208cfe48-b4fb-4bb1-9a00-9dcf15170176

V. Synthetic a priori judgments are contained as principles in all theoretical sciences of reason.

In this passage, Kant is discussing the nature of mathematical judgments and how they are fundamentally synthetic rather than analytic.

It must first be remarked that properly mathematical propositions are always *a priori* judgments and are never empirical, because they carry necessity with them, which cannot be derived from experience

Overall, Kant is challenging the prevailing view that mathematical judgments are analytic and asserting that they are, in fact, synthetic.

- This perspective lays the groundwork for his broader exploration of synthetic a priori judgments in the realms of mathematics and metaphysics in the "Critique of Pure Reason."

More here
https://chat.openai.com/share/208cfe48-b4fb-4bb1-9a00-9dcf15170176

The Transcendental Doctrine of Elements: (B33-B76)

> The Transcendental Doctrine of Elements specifically deals with the conditions under which knowledge is possible and focuses on the nature of *a priori* knowledge.

The Transcendental Doctrine of Elements is divided into two main parts: the Transcendental Aesthetic and the Transcendental Logic.

Transcendental Idealism
Kant argues that space and time do not exist outside of the human mind and experience (they are intuitions of the mind) == the external world does not have space or time

- So if every human seized to exist then so would time and space

Appearances vs Things in Themselves

- When we perceive an object, Kant argues, that we are only experiencing the "appearance of objects"
  - What we cognize is not what actually exists outside of our minds
- As the data of the "things in themselves" goes through the lens of our minds, our intuitions organize this data == we "intuit" them in the words of Kant
  - Our minds orders the raw data into a spatial-temporal format so we can understand it

All objects exist outside of our mind but what we perceive are just the appearance of the objects = The Phenomena

Outside of our minds these objects exist independently (the things in themselves) = The Noumena

thought without content are empty, intuitions without concepts are blind

- Without content, thoughts or concepts lack specific instances or empirical data to give them meaning. In other words, if we have concepts devoid of any real-world examples or content, they remain empty and devoid of significance

- Intuitions, on the other hand, refer to sensory perceptions or raw sensory data that we receive from the external world. According to Kant, these intuitions need to be organized and structured by the mind through the use of concepts.
    - In this sense, intuitions without concepts are "blind" because they lack the ability to contribute meaningfully to our understanding of the world. They remain a chaotic array of sensations without any coherent interpretation.

Kant is emphasizing the interdependence of concepts and intuitions in the cognitive process.

- Concepts provide the structure and meaning to our thoughts, while intuitions supply the raw sensory data that requires conceptual organization.
    - The synthesis of concepts and intuitions is essential for meaningful cognition and the formation of knowledge.

The Transcendental Aesthetic

"I call a science of all principles of *a priori* sensibility the **transcendental aesthetic**."

Space and Time: Kant argues that space and time are not concepts derived from experience but are a priori intuitions necessary for experience. They are the fundamental forms of human sensibility.

- Space and time are not things in themselves but rather frameworks through which we perceive and organize sensory information.
- They provide the necessary conditions for experience, making empirical knowledge possible.
- In my own words: An intrinsic understanding of space and time is presupposed; it is necessary to even carve out experience in the first place. This is why Kant argues that they are *a priori* or **pure intuitions**

Sensibility vs. Understanding: Kant distinguishes between sensibility (the capacity to receive sensory impressions) and understanding (the capacity to think and conceptualize).

- He asserts that knowledge arises from the combination of sensibility and understanding.
- Sensibility gives us intuitions (raw sensory data), and understanding processes and organizes these intuitions through categories (a priori concepts).

The Transcendental Logic

Understanding and Categories: In this part, Kant explores the nature of the understanding and the role of categories in organizing our experiences.

- Categories are a priori concepts that the mind imposes on the raw data of intuition to make sense of it.
    - Kant identifies twelve categories, such as causality, substance, and quantity, which structure our understanding of the world.

Transcendental Deduction: Kant's Transcendental Deduction is a key argument in the Transcendental Logic. It aims to demonstrate that the categories are necessary for the possibility of experience.

- Kant argues that the mind actively imposes its categories on sensory experience, shaping and structuring it into a coherent whole.
    - Without these categories, experience would be chaotic and unintelligible.

Limits of Knowledge: While emphasizing the importance of a priori concepts, Kant also recognizes the limits of human knowledge. He introduces the distinction between phenomena (appearances as we perceive them) and noumena (things in themselves).

- According to Kant, we can only know phenomena, as our knowledge is constrained by the conditions of our perception and understanding.

Hume, A Treatise of Human Nature

> Of the origin of our ideas — Of the idea of necessary connexion + Rules by which to judge of causes and effects

ChatGPT: Hume's exploration of the origin of ideas laid the groundwork for empiricism and influenced later philosophers. His emphasis on the importance of experience, impressions, and the association of ideas challenged traditional philosophical views and had a significant impact on the development of modern philosophy

Of the origin of our ideas

General proposition: That all our simple ideas in their first appearance are derived from simple impressions, which are correspondent to them, and which they exactly represent

The simple impressions always take the precedence of their correspondent ideas but never appear in their contrary order

Simple ideas are not always derived from the correspondent impressions (color example)

- "though the instance is so particular and singular, that it is scarce worth our observing, and does not merit that for it alone we should alter our general maxim."

Impressions and Ideas

- Hume distinguishes between impressions and ideas
  - Impressions are more vivid perceptions we have (sensations/emotions)
  - Ideas are fainter copies of impressions

Simple and Complex Ideas

- Hume categorizes ideas into simple and complex
  - Simple ideas are derived directly from impressions
  - Complex ideas result from combining and compounding simple ideas
- He contends that complex ideas can be broken down into their simple components, which in turn can be traced back to corresponding impressions

Of the idea of necessary connexion

Hume challenges the traditional philosophical notion of causation, which posits a necessary connection between cause and effect

- He argues that when we observe one event followed by another we do not directly perceive any necessary connection between them.
- Instead we only observe a constant conjunction — the regular succession of one event after another

Hume argues that the idea of necessary connection is not derived from sensory experience but is a product of habit and custom

- We become accustomed to seeing certain events follow one another, and our minds form an association between them

- However, this association is a psychological tendency rather than a rational necessity

Hume contends that we cannot prove the existence of a necessary connection between cause and effect through reason alone

- He criticizes the idea that there is a hidden force or power underlying causation, arguing that such a notion goes beyond the evidence provided by our senses

Hume's definition of cause and effect:
Cause 1: An object precedent and contiguous to another, and where all the objects resembling the former are placed in like relations of precedency and contiguity to those objects that resemble the latter

- If this definition be esteemed defective, because drawn from objects foreign to the cause, we may substitute this other definition in its place
  Cause 2: An object precedent and contiguous to another, and so united with it, that the idea, of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other

Rules by which to judge of causes and effects

- Hume is providing guidelines or rules that describe how humans typically form associations between events and come to believe in causal relationships.
- He emphasizes that these rules are based on observed regularities and habits of thought rather than on any inherent necessity in the events themselves.

1. **Constant Conjunction:** Hume argues that we often associate two events as cause and effect because we observe a constant conjunction between them. If we consistently see one event followed by another, we tend to infer a causal relationship. However, Hume emphasizes that this constant conjunction, in itself, does not explain why one event causes another.
2. **Contiguity in Space and Time:** Hume notes that we are more likely to attribute a causal connection between two events if they are spatially and temporally contiguous. Events that occur close to each other in space and time are more likely to be associated as cause and effect.
3. **Priority of Time:** When events are observed in a certain temporal order, with one event consistently preceding the other, we tend to infer a causal relationship. The idea is that the cause must precede its effect.

4. **Distinct Perception:** According to Hume, the cause and effect must be distinctly perceived. If the connection between the cause and effect is not immediately apparent or is hidden from our senses, our ability to establish a causal relationship is weakened.
5. **Logical Connection:** Hume challenges the notion that there is a logical connection or necessity between cause and effect. He argues that even when we observe a constant conjunction and contiguity, the idea of necessary connection is not derived from reason but is a result of custom or habit.

Philosophical Foundations Lecture

#Philosophy

Descartes

- 1st meditation
  - Methodological doubt - a quest for certainty
  - Senses cannot be trusted
  - Ideas can only enter through the senses (e.g. Hume)
  - An evil genius deceiving him
- 2nd Meditation
  - Cogito ergo sum
  - Descartes moves from doubt --> certainty
    - While he has cast doubt on everything, he realizes that the very act of doubt implies a thinking, doubting self—the famous statement "Cogito, ergo sum" (I think, therefore I am)
  - Dualism (mind/body)
    - *res cogitans* vs *res extensa*
    - The mind is more easily inquired than the body

Hume

- Impressions
  - "...sensations, passions and emotions, as they make their first appearance in the soul"
    - Primary
- Ideas
  - "... the faint image of these in thinking and reasoning"
    - derived
- Simple perceptions - both impressions and ideas - cannot be separated further
  - They are given to the mind

- Complex perceptions
  - Can be divided into underlying simple perceptions

Formal Logic

Logic

Modus ponens
P --> Q

Impression	-->	Idea
It rains	thus	The streets are wet
T	T	T
T	F	F
F	T	T
F	T	F

The "antecedent" is the thing appearing before the arrow in the argument

- the first part of a logical proposition

The "consequent" is the thing appearing after the arrow in the argument

- Last part

Hume Again, Cause and Effect

// Section XV

1. The cause and effect must be contiguous in space and time
2. The cause must be prior to the effect
3. There must be a constant union betwixt the cause and the effect
4. The same cause always produces the same effect, and the same effect never arises but from the same cause
5. [...] where several different objects produce the same effect, it must be by means of some quality [...] common amongst them

6. The difference in the effects of two resembling objects must proceed from that particular, in which they differ
7. When any object increases or diminishes with the increase or diminution of its cause, ‘tis to be regarded as a compounded effect, deriv’d from the union of the several different effects, which arise from the several different parts of the cause
8. [...] an object, which exists for any time in its full perfection without any effect, is not the sole cause of that {effect, but requires to be assisted by some other principle, which may forward its influence and operation

The necessity is in the mind not in the objects

Divisions of knowledge

	analytic	**synthetic**
a priori	a triangle has 3 sides = possible	? = necessary
a posteriori	= not really possible	tigers live India = contingent

Critique of Pure Reason, Kant

Epistemological idealism

- knowledge about the world comes primarily from knowledge of the mind

Copernican revolution:

> !! "If intuition has to conform to the constitution of the objects, then I do not see how we can know anything of them a priori, but if the object (as object of the senses) conforms to the constitution of our faculty of intuition, then I can very well represent this possibility to myself"

From object centered to mind centered (object of the senses)

The transcendental aesthetic

- Space & time shape our experience

- however much we abstract away from the particulars about an object, we cannot abstract away that it is placed in space and time

Cognitive Science

- finding out what appears given to the mind, may reveal facts about the mind rather than about objects

MANIFEST IMAGE AND SCIENTIFIC IMAGE

- Similar to distinction between things-as-they-appear-to-us/things-in-themselves?

Philosophical Foundations

#Philosophy

Things to look out for

- Descartes
  - Find the *cogito sum*
- Hume
  - What is meant by a constant conjunction?
- Kant
  - Make sure you understand his Copernican revolution
  - And this slogan: *thought without content are empty, intuitions without concepts are blind*
- Sellars
  - How does the scientific image and the manifest image relate to Kant's philosophy

Descartes - Meditations I and II

Hume, A Treatise of Human Nature

Kant, Critique of Pure Reason

Sellars, Philosophy and the Scientific Image of Man

Sellars, Philosophy and the Scientific Image of Man

> One Sellars' most influential ideas is the distinction between the "manifest image" and the "scientific image" of man.

Manifest Image:

- The manifest image refers to the way we naturally perceive and understand the world in everyday life. It encompasses our common-sense beliefs, language, and the conceptual framework we use to make sense of our experiences.
- This image is shaped by ordinary language and is grounded in our immediate, pre-philosophical experiences. It includes concepts like colors, sounds, shapes, and mental states as we commonly understand them.

> "It is the framework in terms of which, to use an existentialist turn of phrase, man first encountered himself"

Sellars on the transition from preconceptual behavior to conceptual thinking (from no consciousness to consciousness, I'm guessing)

> The conclusion is difficult to avoid that the transition from preconceptual patterns of behavior to conceptual thinking was a holistic one, a jump to a level of awareness which is irreducibly new, a jump which was the coming into being of man

On what basic objects are included in the manifest image:

> "when we say that it includes persons, animals, lower forms of life, and "merely material" things, like rivers and stones. The list is not intended to be

complete, although it is intended to echo the lower stages of the "great chain of being" of the platonic tradition."

On categories in the manifest image:

> "...the world can be construed as progressive prunings of categories pertaining to the person and his relation to other persons and the group"

Scientific Image:

- The scientific image, on the other hand, is the image of the world as described by the scientific theories of the day. It involves scientific concepts and explanations that often go beyond our common-sense understanding.
- Scientific images are formulated using precise, technical language and mathematical representations. They may include abstract entities like quarks, genes, or electromagnetic fields that are not part of our everyday experiences.

Sellars on the distinction between the manifest image and the scientific image

> "It will be remembered that the contrast I have in mind is not that between an unscientific conception of man-in-the-world and a scientific one, but between that conception which limits itself to what correlational techniques can tell us about perceptible and introspectible events (manifest, red) and that which postulates imperceptible objects and events for the purpose of explaining correlations among perceptibles (scientific, red)"

And again..:

> "Our contrast, then, is between two ideal constructs: **(a)** the correlational and categorial refinement of the "original image," which refinement I am calling the manifest image; **(b)** the image derived from the fruits of postulational theory construction which I am calling the scientific image."

How the two images blend together:

Sellars on the nature of these images

> "they are both "idealizations" in something like the sense in which a frictionless body or an ideal gas is an idealization"

Sellars often employs a metaphor of the "stereoscopic view" i.e. how these two distinct images blend together

Conclusions

Sellars was interested in exploring the relationship between these two images and understanding how they relate to each other. He argued that both images are legitimate and necessary for a comprehensive understanding of reality, but they operate in different spheres:

The Autonomy of the Scientific Image:

- Sellars believed in the autonomy of the scientific image. He argued that scientific explanations provide a deeper and more accurate understanding of the world than the manifest image. While the manifest image remains valid in our everyday lives, it is incomplete and, in some cases, even inconsistent when compared to the scientific image.

The Irreducibility of the Manifest Image:

- Despite the importance of the scientific image, Sellars acknowledged the irreducibility of the manifest image. Common-sense concepts are indispensable for understanding human experience and organizing our practical lives, even though they may not capture the ultimate structure of reality.

Conceptual Change and Evolution: Sellars is interested in how our conceptual frameworks evolve over time, especially in response to scientific advancements.

- He recognizes that as scientific theories develop, our language and conceptual schemes must also adapt to accommodate these changes. This challenges traditional epistemological notions of stable, ahistorical concepts.

How does the scientific image and the manifest image relate to Kant's philosophy?

- While there are parallels, it's important to note that Sellars' philosophy represents a departure from some aspects of Kantian thought. For instance,
  - Sellars is more optimistic about our ability to attain knowledge about the external world through scientific inquiry, whereas Kant's epistemology is more limiting in terms of what we can know about things in themselves (the noumena).

Bem - Feeling the Future Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

#Philosophy

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms.

- The term is purely descriptive; it neither implies that such phenomena are paranormal nor connotes anything about their underlying mechanisms.
- Alleged psi phenomena include
  - telepathy, the apparent transfer of information from one person to another without the mediation of any known channel of sensory communication;
  - clairvoyance (sometimes called remote viewing), the apparent perception of objects or events that do not provide a stimulus to the known senses;
  - psychokinesis, the apparent influence of thoughts or intentions on physical or biological processes;
  - and precognition (conscious cognitive awareness) or premonition (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process.

Individual differences. There were no significant sex differences in the present experiment. Over the years, however, the trait of extraversion has been frequently reported as a correlate of psi, with extraverts achieving higher psi scores than introverts

- A meta-analysis of 60 independent experiments published between 1945 and 1983, involving several kinds of psi tasks, revealed a small but reliable correlation between extraversion and psi performance'

Over the course of our research program—and within the experiment just reported—we have obtained positive results using both PRNGs and a true RNG, arguably leaving

precognition/ reversed causality the only nonartifactual interpretation that can account for all the positive results

The article covers 9 different experiments which is really condensed reading so I skipped to the general discussion part:

General Discussion

Table 7.png summarizes the significance levels and effect sizes obtained in the nine experiments reported in this article. As seen in the table, the mean effect size across all the experiments was .22, and all but the retroactive induction of boredom experiment yielded statistically significant results. Stimulus seeking was significantly correlated with psi performance in five of the experiments

One of the major obstacles to successful replication in psychology generally is the influence of the experimenter on the results. The sex, age, and demeanor of the experimenter can interact with characteristics of the participants, and expectancies of the experimenter can affect the results in subtle ways (Rosenthal, 1966)

If psi exists, then it is not unreasonable to suppose that it might have been acquired through evolution by conferring survival and reproductive advantage on the species

Familiarity and belief: But how can it be like that? Among psi phenomena, precognition and retroactive influence might seem to be the most anomalous because they challenge not only our classical conceptions of space and distance, as telepathy and clairvoyance do, but also those of time and causality.

- Although less well known than discussions of nonlocality, alternative conceptions of time and causality also constitute an active area of discussion within physics (Barbour, 2001)

Why is Research into Precognition Seen as Dubious?

Research into precognition, the ability to perceive or predict future events before they happen, is often viewed with skepticism and considered dubious by the scientific community for several reasons

1. **Lack of Reproducibility:** Many studies claiming to demonstrate precognition have not been successfully replicated by independent researchers. Reproducibility is a crucial aspect of scientific research, and when experiments

cannot be consistently reproduced, it raises doubts about the validity of the findings.

2. **Methodological Issues:** Some studies investigating precognition have been criticized for methodological flaws, such as small sample sizes, inadequate controls, and the absence of rigorous experimental designs. These issues can compromise the reliability and validity of the results.
3. **Publication Bias:** Positive results are more likely to be published than negative ones, creating a bias in the available literature. This publication bias can contribute to an overestimation of the evidence supporting precognition and make the phenomenon appear more credible than it might be. (what the article calls "file drawer effect")
4. **Lack of Theoretical Framework:** Precognition lacks a well-established theoretical framework within the broader scientific understanding of physics, biology, or psychology. The absence of a solid theoretical foundation makes it challenging for researchers to design experiments that align with accepted scientific principles.
5. **Extraordinary Claims Require Extraordinary Evidence:** The principle that "extraordinary claims require extraordinary evidence" is a common guideline in science. Given that precognition challenges our understanding of causality and the nature of time, the evidence supporting it would need to be exceptionally robust to gain widespread acceptance.
6. **Psi Controversy:** Precognition is often associated with the broader field of psi phenomena, which includes telepathy and clairvoyance. The study of psi has been controversial, and many scientists remain skeptical due to a perceived lack of solid empirical evidence.

Hansson - Science and Pseudo-Science

#Philosophy

The Purpose of Demarcations

Demarcations of science from pseudoscience can be made for both theoretical and practical reasons (Mahner 2007, 516). From a theoretical point of view, the demarcation issue is an illuminating perspective that contributes to the philosophy of science in much the same way that the study of fallacies contributes to our knowledge of informal logic and rational argumentation. From a practical point of view, the distinction is important for decision guidance in both private and public life

Goes on to give a bunch of examples where it is important (climate, healthcare, journalism, whathaveya)

The “science” of pseudoscience

The ultimate issue is “how to determine which beliefs are epistemically warranted” (Fuller 1985, 331). In a wider approach, the sciences are fact-finding practices, i.e., human practices aimed at finding out, as far as possible, how things really are (Hansson 2018).

- Other examples of fact-finding practices in modern societies are journalism, criminal investigations, and the methods used by mechanics to search for the defect in a malfunctioning machine.

In this perspective, the demarcation of science is a special case of the delimitation of accurate fact-finding practices.

- The delimitation between science and pseudoscience has much in common with other delimitations, such as that between accurate and inaccurate journalism and between properly and improperly performed criminal investigations (Hansson 2018).

The “pseudo” of pseudoscience

All non-science is not pseudoscience, and science has nontrivial borders to other non-scientific phenomena, such as metaphysics, religion, and various types of non-scientific systematized knowledge. (Mahner (2007, 548) proposed the term “parascience” to cover non-scientific practices that are not pseudoscientific.) Science also has the internal demarcation problem of distinguishing between good and bad science.

A comparison of the negated terms related to science can contribute to clarifying the conceptual distinctions. “Unscientific” is a narrower concept than “non-scientific” (not scientific), since the former but not the latter term implies some form of contradiction or conflict with science.
“Pseudoscientific” is in its turn a narrower concept than “unscientific”. The latter term differs from the former in covering inadvertent mismeasurements and miscalculations and other forms of bad science performed by scientists who are recognized as trying but failing to produce good science.

==Etymologically "Pseudo" means false ==
In accordance with this, the Oxford English Dictionary (OED) defines pseudoscience as follows:

- “A pretended or spurious science; a collection of related beliefs about the world mistakenly regarded as being based on scientific method or as having the status that scientific truths now have.”

Pseudoscience is defined (by some) as

1. Not scientific
2. Its major proponents try to create the impression that it is scientific

An immediate problem with the definition based on (1) and (2) is that it is too wide. There are phenomena that satisfy both criteria but are not commonly called pseudoscientific. One of the clearest examples of this is fraud in science.

The term “science” has both an individuated and an unindividuated sense.

- In the individuated sense, biochemistry and astronomy are different sciences, one of which includes studies of muscle proteins and the other studies of supernovae. The Oxford English Dictionary (OED) defines this sense of science as “a particular branch of knowledge or study; a recognized department of learning”.
- In the unindividuated sense, the study of muscle proteins and that of supernovae are parts of “one and the same” science. In the words of the OED, unindividuated science is “the kind of knowledge or intellectual activity of which the various ‘sciences‘ are examples”.

Pseudoscience is an antithesis of science in the individuated rather than the unindividuated sense. There is no unified corpus of pseudoscience corresponding to the corpus of science. For a phenomenon to be pseudoscientific, it must belong to one or the other of the particular pseudosciences.

The objects of demarcation

Elliot Sober and Creationism

#Philosophy

Elliot Sober - Creationism

Emphasizes the importance of avoiding present-day perspectives when evaluating historical ideas (phrenology as an example).

A person can behave pigheadedly toward propositions that are perfectly scientific.

- Flat earth theory is a scientific proposition
- It can be tested by scientific means, which is why we are entitled to regard it as false.
- Yet, flat-earthers are not behaving scientifically when they dogmatically accept this perfectly testable proposition even though there is lots of evidence against it.

Creationists maintain that some characteristics of living things are the result of intelligent design by God; they deny that natural processes suffice to account for all features of living things.

- Is creationism a scientific theory? If so, why do scientists fail to take it seriously?

Genuinely scientific theories are extended and refined over time in ways that allow new observations to be brought to bear.

- The intellectual stagnation that one finds in creationist thought is a sign that something has gone wrong

Before Darwin's time, some of the best and the brightest in both philosophy and science argued that the adaptedness of organisms can be explained only by the hypothesis that organisms are the product of intelligent design.

- This line of reasoning--the *design argument*-- is worth considering as an object of real intellectual beauty

Paley's Watch and the Likelihood Principle

In the *Summa Theologiae*, St. Thomas Aquinas (1224-1274) presented five ways to prove that God exists.

- The fifth of these was the "argument from design." Aquinas's version of the design argument elaborated ideas already put forward by Plato and Aristotle

Many philosophers now regard David Hume *Dialogues Concerning Natural Religion* (1779) as the watershed in this argument's career.

- Before Hume, it was possible for serious people to be persuaded by the argument, but after the onslaught of Hume's corrosive skepticism, the argument

was in shambles and has remained that way ever since.

William Paley in his *Natural Theology* (1805).

- Organisms are intricate and well adapted. Their complexity is not a jumble of uncoordinated parts; rather, when we examine the parts with the utmost care, we discern how the different parts contribute to the wellfunctioning of the organism as a whole
- Paley considers two possible explanations of these observations

1. organisms were created by an intelligent designer. God is an engineer who built organisms so that they would be well suited to the life tasks they face
2. random physical forces acted on lumps of matter and turned them into living things. Paley's goal is to show that the first explanation is far more plausible than the second

Paley constructed an analogy:
Suppose you were walking across a heath and found a watch. You open the back of the watch and observe that it is intricate and that its parts are connected in such a way that the watch as a whole is well suited to the task of timekeeping. How might you explain the existence and characteristics of this object?

1. Intelligent design
2. Random physical processes produced the watch

- Paley is a fan of the intelligent design argument

Dear Elliot gets into some probability theory here. On the topic of why there is a difference between $P(O|H)$ and $P(H|O)$, which can be hard to comprehend fully, he gives this fine example:

- Let O be the observation statement, "there is rumbling in the attic."
- Let H be the hypothesis, "there are gremlins in the attic, and they are bowling."
- I hope you see that, $P(O|H)$ is very high but $P(H|O)$ is not high at all.
    - If there actually were gremlins bowling up there, we would expect to hear noise.
    - But the mere fact that we hear the noise does not make it very probable that there are gremlins bowling.
- The gremlin hypothesis has a high likelihood but a low probability, given the noises we hear.

Returning to Paleys arguments, we can represent the statements involved in the watch argument as follows:
A: The watch is intricate and well suited to the task of timekeeping
W_1 : The watch is the product of intelligent design
W_2 : The watch is the product of random physical processes

Paley claims that P(A | W 1) > > P(A | W 2). He then says that the same pattern of analysis applies to the following triplet of statements:
B: Living things are intricate and well-suited to the tasks of surviving and reproducing.
L_1 : Living things are the product of intelligent design.
L_2 : Living things are the product of random physical processes.

Paley argues that if you agree with him about the watch, you also should agree that P(B | L 1) > > P(B | L 2). Although the subject matters of the two arguments are different, their logic is the same. Both are inferences to the best explanation in which the Likelihood Principle is used to determine which hypothesis is better supported by the observations.

Hume's Critique

In Hume's view, the argument is not an inference to the best explanation; rather, it is an argument from analogy, or an inductive argument. T

- his alternate conception of the argument makes a great deal of difference.
  - Hume's criticisms are quite powerful if the argument has the character he attributes to it.
  - But if the argument is, as I maintain, an inference to the best explanation, Hume's criticisms entirely lose their bite.

For Hume, this argument rested on an analogy between living things and artifacts:

- Watches are the product of intelligent design.
- Watches and organisms are similar.
- Organisms are the product of intelligent design.

Hume believes that this fairly plausible theory about the logic of analogy arguments has important consequences for the design argument. To see how strongly the premises support the conclusion of the design argument, we must ask how similar watches and organisms really are. A moment's reflection shows that they are very dissimilar

- The immediate consequence, of course, is that the design argument is a very weak analogy argument.
- It is preposterous to infer that organisms have a given property simply because watches happen to have it

Although Hume's criticism is devastating if the design argument is an argument from analogy, I see no reason why the design argument must be construed in this way. Paleys argument about organisms stands on its own, regardless of whether watches and organisms happen to be similar. The point of talking about watches is to help the reader see that the argument about organisms is compelling.

According to Sober, likelihood stands on its own; analogy is irrelevant

Poison ivy, rash argument follows --> Sample size is crucial in Hume's argumentation

- Hume contends that if we are to have good reason to think that the organisms in our world are the product of intelligent design, then we must have looked at lots of other worlds and observed intelligent designers producing organisms there.
- But how many such worlds have we observed? The answer is--not even one. The inductive argument is as weak as it possibly could be; its sample size is zero

Once again, it is important to see that an inference to the best explanation need not obey the rules that Hume stipulates. Consider the suggestion by Alvarez et al (1980) that the mass extinction that occurred at the end of the Cretaceous period was caused by a large meteorite crashing to earth and sending up a giant dust cloud. (...) It is quite irrelevant that we have never witnessed meteorite strikes producing mass extinctions "in other worlds."

- Inference to the best explanation is different from an inductive sampling argument.

Why Natural Selection Isn't a Random Process

Natural selection involves unequal probabilities, i.e. it is not random

Creationists sometimes describe natural selection as "random" when they compare it to a tornado blowing through a junkyard. The tornado "randomly" rearranges the pieces of junk.

- It is enormously improbable that this "random" activity will put together a functioning automobile. Creationists think the same is true of natural selection: Because it is "random," it cannot create order from disorder.

The process of natural selection has two components. First, variation must arise in the population; then, once that variation is in place, natural selection can go to work

Variation is generated without regard to whether it "matches the target" (i.e., is advantageous to the organism). But retention (selection among the variants that arise) is another matter. Some variants have greater staying power than others

- Variation is generated at random, but selection among variants is nonrandom.

According to the theory of natural selection, the organisms in a population retain a characteristic because that characteristic helps them survive and reproduce. It doesn't take an intelligent designer to make some traits advantageous and others deleterious.

Two Kinds of Similarity

The evolutionist argument

Paley stressed the adaptive perfection of nature

- Darwin began the break with this perfectionist tradition, and modern evolutionists have followed Darwin's lead.
- They reject the idea that adaptation is perfect, arguing rather that typically it is **imperfect**
  - What natural selection predicts is that the fittest of the traits actually represented in a population will become common.
  - The result is not the best of all conceivable worlds but the best of the variants actually available. And the word "best" means best for the organism's reproductive success.

Natural selection is a "tinkerer" (Jacob 1977). Consider the fact that organisms in various species often exhibit structural differences among parts that perform the same function.

- Wings in birds, bats, and insects all facilitate flight. Yet, close attention to these "wings" reveals that they differ in numerous respects that have little or nothing to do with the requirements of flight.
- If wings were designed by an intelligent engineer so that they would optimally adapt the organism for flight, it would be very hard to explain these differences.
  - On the other hand, they become readily intelligible if one accepts the hypothesis that each of these groups is descended from wingless ancestors.

- The bird's wing is similar to the forelimbs of its wingless ancestors. A bat's wing is likewise similar to the forelimb of its wingless ancestors.
- Wings were not designed from scratch but are modifications of structures found in ancestors. Because natural selection is a tinkerer, organisms retain characteristics that reveal their ancestry.

A similar argument is based on vestigial organs

- These vestigial traits are entirely useless to the organism. It is puzzling why an intelligent designer would have inserted them into the developmental sequence only to delete them a short time later
- Makes better sense when viewed through the lens of evolution

These arguments make use of the Likelihood Principle

- Some observation (O) is cited, and the design hypothesis (D) and the hypothesis of evolution by natural selection (E) are considered in its light. The claim is made that the observation would be very surprising if the design hypothesis were true but would be quite unsurprising if the hypothesis of evolution by natural selection were correct.
- The observation strongly favors evolution over design because P (O | E) > > P(O | D).

I have formulated the tree of life hypothesis as a very strong (i.e., logically ambitious) claim. It says that all present-day organisms on earth are genealogically related to each other. One standard line of evidence used to answer this question is the (near) universality of the genetic code

- With some minor exceptions, all living things use the same code. This is interpreted as evidence that all terrestrial life is related.

Arbitrary similarity, not adaptive similarity, provides powerful evidence of genealogical relationship.

The Problem of Predictive Equivalence

One possible modification involves removing God from the problem of the origin of species. Suppose one believed that God created the universe and then sat back and let physical laws play themselves out. This version of theism does not conflict with the hypothesis of evolution by natural selection

Consider the hypothesis that God created each species separately but did so in a way that misleads us into thinking that species evolved by natural selection. This hypothesis of a "trickster" God disagrees with the hypothesis of evolution by natural selection

Is the Design Hypothesis Unscientific?

It has been argued that creationism is not a scientific hypothesis because it is untestable.
It should be clear that this line of criticism is not compatible with the likelihood arguments we have reviewed. If creationism cannot be tested, then what was one doing when one emphasized the imperfection of nature? Surely it is not possible to test and find wanting a hypothesis that is in fact untestable

Popper's basic idea is that scientific ideas are falsifiable; they "stick their necks out," whereas unscientific ideas do not.

- Less metaphorically, scientific propositions make predictions that can be checked observationally. They make claims about the world that, at least in principle, are capable of conflicting with what we observe.
- Unscientific claims, on the other hand, are compatible with all possible observations. No matter what we observe, we can always retain our belief in an unscientific proposition.
- According to Popper, our beliefs should be falsifiable, not false.

Popper's criterion of falsifiability requires that we be able to single out a special class of sentences and call them observation sentences. A proposition is then said to be falsifiable precisely when it is related to observation sentences in a special way:

- Proposition P is falsifiable if and only if P deductively implies at least one observation sentence O.
- Falsifiable propositions make predictions about what can be checked observationally; this idea is made precise by the idea that there is a deductive implication relation between the proposition P and some observational report O.

Sober thinks there are a lot of problems with Popper's criterion of falsifiability

Popper (1959, 1963) held that there is an asymmetry between falsification and verification. **He maintained that it is possible to prove theories false but impossible to prove them true**

...

There is a whole lot more, but frankly it does not make sense to me at the current point in time

The Incompleteness of Science

Scientists do not think they now have all the answers. That is why they continue to do research. On the other hand, creationists have at hand an all-purpose explanation for any observation you please. The origin of life, the distribution of modes of reproduction, and everything else can be explained by a four-word hypothesis: "It was God's will."

Pseudoscience Lecture

#Philosophy

Loose definitions

Scientific statements

- About the natural world
- Used to explain how the world works
- Are at least in principle testable
- Science can become pseudoscience (phrenology)

Pseudoscientific statements

- Tries to look like science
- The explanation given supports some agenda
- Are not testable, or do not recognize refuting statements
- Can pseudoscience become science?

Characteristics of pseudoscience

- Belief in authority
- Unrepeatable experiments
- Hand picked examples
- Unwillingness to test
- Disregard of refuting information
- Built in subterfuge
- Explanations are abandoned without explanation

METHINKSITISAWEASEL

- true randomness only has variation, evolution adds the feature of retention
  - $p_{random} = 1/26^{19}$
  - $p_{evolution}$= 474,36 / 10.000

Popper's Asymmetry

In plain words: particular observations cannot verify general theories, but particular observations can falsify general theories

Falsification	*Verification*
If *T*, then *O*	If *T*, then *O*
not-*O*	*O*
not-*T*	*T*
deductively valid	deductively invalid

- If it rains, the streets are wet
- The streets are wet --> not necessarily because it has rained

science can tell us only that a theory is false (or that it has yet to be refuted)

Duhem's thesis:
Typically, T does not deductively imply O; rather, it is T&A that deductively implies O (here, T is a theory, O is an observation statement, and A is a set of auxiliary assumptions).

- This idea is sometimes called Duhem's Thesis; it is named for Pierre Duhem (1861-1916), a physicist, historian, and philosopher of science who noted this pervasive pattern in physical theories.

Science, non-science and pseudoscience

#Philosophy

Things to look out for

- The demarcation problem: how do we differentiate science from non-science and especially how do we differentiate it from pseudoscience?
- E.g.: why is intelligent design not a serious rival of evolutionary theory?
- E.g.: why is research into precognition seen as dubious?

Elliot Sober and Creationism

Bem - Feeling the Future Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Hansson - Science and Pseudo-Science

Carnap - Psychology in Physical Language

#Philosophy

Introduction. Physical Language and Protocol Language

Opens quite strongly:

> "IN WHAT FOLLOWS, we intend to explain and to establish the thesis that every sentence of psychology may be formulated in Physical language. To express this in the material mode of speech: all sentences of psychology describe physical occurrences, namely, the physical behavior of humans and other animals."

and

> "This is a sub-thesis of the general thesis of *physicalism* to the effect that physical language is a *universal language*, that is, a language into which every sentence may be translated"

If our thesis is correct, the generalized sentences of psychology, the laws of psychology, are also translatable into physical language. They are thus physical laws.

The Forms of Psychological Sentences

- A singular psychological sentence e.g. “Mr. A was angry at noon yesterday” (an analogue of the physical sentence, “Yesterday at noon the temperature of the air in Vienna was 28 degrees centigrade”), is concerned with a particular person at a particular time
- General psychological sentences have various forms, of which the following two are perhaps the most important
    1. A sentence may describe a specific quality of a specific kind of event, e.g. “An experience of surprise always (or: always for Mr. A, or: always for people of such and such a society) has such and such a structure.” A physical analogy would be: “Chalk (or: chalk of such and such a sort) always is white.”
    2. universal-conditional statements concerning sequences of events, that is, of causal laws. For instance,
        1. “When, under such and such circumstances, images of such and such a sort occur to a person (or: to Mr. A, or: to anyone of such and such a society), an emotion of such and such a sort always (or: frequently, or: sometimes) is aroused.” A physical analogy would be: “When a solid body is heated, it usually expands.”

Phenomenology claims to be able to establish universal synthetic sentences which have not been obtained through induction. These sentences about psychological qualities are, allegedly, known either *a priori* or on the basis of some single illustrative case

- In our view, knowledge cannot be gained by such means. We need not, however, enter upon a discussion of this issue here, since even on the view of phenomenology itself, these sentences do not belong to the domain of psychology.

In physics it sometimes seems to be the case that a general law is established on the basis of some single event. For instance,

- if a physicist can determine a certain physical constant, say, the heat-conductivity of a sample of some pure metal, in a single experiment, he will be convinced that, on other occasions, not only the sample examined but any similar sample of the same substance will, very probably, be characterizable by the same constant.
- But here too induction is applied. As a result of many previous observations the physicist is in possession of a universal sentence of a higher order which enables him in this case to follow an abbreviated method. This higher-order sentence reads roughly: “All (or: the following) physical constants of metals vary only slightly in time and from sample to sample.”

The situation is analogous for certain conclusions drawn in psychology. If a psychologist has, as a result of some single experiment, determined that the simultaneous sounding of two specific notes is experienced as a dissonance by some specific person A, he infers (under favorable circumstances) the truth of the general sentence which states that the same experiment with A will, at other times, have the same result

- Here too the inference from a singular sentence to a general one is only apparent. Actually, a sentence inductively obtained from many observations is brought into service here

It thus remains the case that every general sentence is inductively established on the basis of a number of singular ones.

Finally, we must consider sentences about psycho-physical interrelations, such as for instance, the connection between physical stimulus and perception. These are likewise arrived at through induction, in this case through induction in part from physical and in part from psychological singular sentences. **The most important sentences of gestalt psychology belong also to this kind.**

General sentences have the character of hypotheses in relation to concrete sentences, that is, the testing of a general sentence consists in testing the concrete sentences which are deducible from it. A general sentence has content insofar and only insofar as the concrete sentences deducible from it have content. Logical analysis must therefore primarily be directed towards the examination of the latter sort of sentences

Sentences About Other Minds

The epistemological character of a singular sentence about other minds will now be clarified by means of an analogy with a sentence about a physical property, defined as a disposition to behave (or respond) in a specific manner under specific circumstances (or stimuli).

A Sentence about a property of a physical substance

Example: I assert the sentence P_1: “This wooden support is very firm.”

A Sentence about a condition of some other mind.

Example: I assert the sentence P_1: "Mr. A is now excited.”

These are two different ways in which sentence P_1 may be derived. We shall designate them as the “ rational ” and the “ intuitive ” methods

The rational method consists of inferring P_1 from some protocol sentence p_1 (or from several like it), more specifically, from a perception-sentence

- about the shape and color of the wooden support.
- about the behavior of Mr. A, e.g. about his facial expressions, his gestures, etc., or about physical effects of A's behavior, e.g. about characteristics of his handwriting

In order to justify the conclusion, a major premise O is still required, namely the general sentence which asserts that

- when I perceive a wooden support to be of this color and form, it (usually) turns out to be firm. (A sentence about the perceptual signs of firmness.)
- when I perceive a person to have this facial expression and handwriting he (usually) turns out to be excited. (A sentence about the expressional or graphological signs of excitement.)

Lau says:

- $P_1 \Leftrightarrow (p_1 \wedge O)$
- "P1 is true if and only if both p1 and O are true simultaneously." It's expressing an equivalence relationship between P1 and the conjunction of p1 and O.
    - Creds to ChatGPT

The content of P_1 does not coincide with that of p_1, but goes beyond it. This is evident from the fact that to infer P_1 from p_1 O is required. The cited relationship between P_1 and p_1 may also be seen in the fact that under certain circumstances, the inference from p_1 to P_1 may go astray. It may happen that, though p_1 occurs in a protocol, I am obliged, on the grounds of further protocols, to retract the established system sentence P_1. I would then say something like, "I made a mistake. The test has shown:

- that the support was not firm, even though it had such and such a form and color."
- that A was not excited, even though his face had such and such an expression."

In practical matters the intuitive method is applied more frequently than this rational one, which presupposes theoretical knowledge and requires reflection.
In accordance with the intuitive method, P_1 is obtained without the mediation of any other sentence from the identically sounding protocol sentence p_2.

- "The support is firm."
- "A is excited."

Consequently, one speaks in this case of *immediate perceptions*

- of properties of substances, e.g., of the firmness of supports.

- of other minds, e.g., of the excitement of A

Here too we can best clarify the difference by considering the possibility of error. It may happen that, though p_2 occurs in my protocol, I am obliged, on the basis of further protocols, to retract the established system sentence P_1. I would then say "I made a mistake. Further tests have shown:

- that the support was not firm, although I had the intuitive impression that it was."
- that A was not excited, although I had the intuitive impression that he was."

Bunch of text wont write, so picture:

respective fields, the majority of them nowadays would give us thoroughly non-analogous answers. The identity of the content of P_2

and of the content of the physical sentence P_1 would be agreed to as a matter of course by all physicists.

and of the content of the psychological sentence P_1 would be denied by almost all psychologists (the exceptions being the radical behaviorists).

The contrary view which is most frequently advocated by psychologists is that, "A sentence of the form of P_1 asserts the existence of a state of affairs not identical with the corresponding physical structure, but rather, only accompanied by it, or expressed by it. In our example:

P_1 states that the support not only has the physical structure described by P_2, but that, besides, there exists in it a certain force, namely its *firmness*.

This firmness is not identical with the physical structure, but stands in some parallel relation to it in such a manner that the firmness exists when and only when a physical structure of the characterized sort exists.

Because of this parallelism one may consider the described reaction to certain stimuli—which is causally dependent upon that structure—to be an *expression* of firmness.
Firmness is thus an occult property, an obscure power which stands behind physical structure, appears in it, but itself remains unknowable."

P_1 states that Mr. A not only has a body whose physical structure (at the time in question) is described by P_2, but that—since he is a *psychophysical being*—he has, besides, a consciousness, a certain power or entity, in which that excitement is to be found.
This excitement cannot, consequently, be identical with the cited structure of the body, but stands in some parallel relation (or in some relation of interaction) to it in such a manner that the excitement exists when and only when (or at least, frequently when) a physical, bodily structure of the characterized sort exists.
Because of this parallelism one may consider the described reaction to certain stimuli to be an *expression* of excitement.

Excitement, or the consciousness of which it is an attribute, is thus an occult property, an obscure power which stands behind physical structure, appears in it, but itself remains unknowable."

P_2: a dispositional statement (logical behaviorism)

This view falls into the error of a hypostatization[1] as a result of which a remarkable duplication occurs: besides or behind a state of affairs whose existence is empirically determinable, another, parallel entity is assumed, whose existence is not determinable. (Note that we are here concerned with a sentence about other minds.) But—one may now object—is there not really at least one possibility of testing this

claim, namely, by means of the protocol sentence p2 about the intuitive impression of the firmness of the support?

The objector will point out that this sentence, after all, occurs in the protocol along with the perception sentence p1. May not then a system sentence whose content goes beyond that of P2 be founded on p2? This may be answered as follows.

- A sentence says no more than what is testable about it. If, now, the testing of Pl consisted in the deduction of the protocol sentence p2, these two sentences would have the same content. But we have already seen that this is impossible

There is no other possibility of testing Pl except by means of protocol sentences like p1 of p2. If, now, the content of Pl goes beyond that of P2, the component not shared by the two sentences is not testable, and is therefore meaningless. If one rejects the interpretation of Pl in terms of P2, Pl becomes a metaphysical pseudo-sentence.

1. Wtf↩

Logical Positivism Lecture

#Philosophy

Learning Goals

- Understanding how logical positivism separates meaningful from meaningless sentences
- Understanding the implications for philosophy, science and especially, psychology
- Understanding the problems facing logical positivism, e.g. verification of general sentences and the Quine-Duhem thesis

3 Kinds of meaningful sentences

True by virtue of their form
Examples:

- "It is raining or it is not raining" has the form

    - A v ¬A (a tautology)
- A bachelor is an unmarried man (analytic)
    - The predicate "unmarried man" is included in the subject "bachelor"
- 2 + 2 = 4
    - Follows the rules of mathematics

A	v	¬A
T	T	FT
F	T	TF

False by virtue of their form

Examples:

- "It is raining and it is not raining" has the form
    - A ∧ ¬A (a contraction)
- 2 + 2 = 5
    - Does not follow the rules of mathematics

Empirical statements

"With respect to all other statements the decision about truth and falsehood lies in the protocol sentences" Carnap, p. 76, EoM

Protocol sentences = observation sentences (*O*)

- Examples
    - The liquid became transparent at time *t*
    - The magnetoencephalogram spiked across all sensors at time *t*
    - The measurement bar extended to *x* cm

requirement: *O*'s need to be expressed in a language that is *universal* and *intersubjective*

What about all the sentences of our language that are not O's?

Non-O's

lets call them (T)heoretical statements

- Examples

- There exists a gravitational force: $F = \frac{G \cdot m_1 \cdot m_2}{r^2}$
- Automatic processes of the mind conflict with controlled processes of the mind: Stroop task (green red)
- The principle of the world is water (attributed to Thales)
- The thing-in-itself lies beyond the world of appearance (Kant)

What sentences are deducible from *T*?
Criterion of meaning:
a T-statement is only meaningful if (an) O statement(s) follow from it
or from an Epistemological View:

- How is T to be verified?

Thus if we cannot specify, i.e. deduce O from T:
T → O
then T is meaningless

What are pseudostatements?
Statements with pseudowords in them

- teavy: no criterion of application
- toovy: really just a synonym for quadrangular
- principle: in the metaphysical sense

Thesis of epistemological physicalism:

- “The physical language is universal and intersubjective”
- “... there would, basically, be only one kind of object—physical occurrences ...”
- “... psychology [...] may formulate its sentences as it pleases—these sentences will [...] be translatable into physical language”

Singular and General sentences

- Singular psychological sentence
  - “Mr. A was angry at noon yesterday”
- Singular physical sentence –
  - “Yesterday at noon the temperature of the air in Vienna was 28 degrees centigrade”

- General psychological sentence (type I)
  - "An experience of surprise always [...] has such and such a structure"
- General physical sentence (type I)
  - “Chalk [...] always is white”
- General psychological sentence (type II)
  - “When, under such and such circumstances, images of such and such a sort occur to a person [...], an emotion of such and such a sort always [...] is aroused”
- General physical sentence (type II)
  - "When a solid body is heated, it usually expands"

$P_1 \Leftrightarrow (p_1 \wedge O)$

Challenges to logical positivism

Analytic/synthetic distinction not clear (dogma 1)

- There is no good definition of analyticity
  - Definition, interchangeability, semantical rules all fail
    - And thus the analytic/synthetic distinction cannot be upheld
- The idea that T1, T2, Tn-1, Tn

"The total field is so *underdetermined* by its boundary conditions, *experience*, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience" Quine, p 60

"Any statement can be held true come what may, if we take drastic enough adjustments elsewhere in the system" Quine, p 60

Pragmatism
"The myth of physical objects is *epistemologically superior* to mist in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience"

The raven paradox
R: is a raven
B: is black

$R \rightarrow B \vdash \neg B \rightarrow R$

Induction problem

Quine - Two Dogmas of Empiricism

#Philosophy

W.V. Quine's "Two Dogmas of Empiricism," published in 1951, challenges the traditional distinction between analytic and synthetic statements and the verification theory of meaning.

In the essay, Quine argues against the notion that statements can be divided into those that are true by definition (analytic) and those that are empirically verifiable (synthetic). He contends that this distinction is untenable and suggests a holistic view of language, where sentences gain meaning within a web of interconnected beliefs.

Quine's critique of the analytic-synthetic distinction and his emphasis on the interdependence of our beliefs have had a profound impact on the philosophy of language and epistemology.

Essay bullet points:

- **Analytic-Synthetic Distinction Critique:**
  - Challenges the traditional division of statements into analytic (true by definition) and synthetic (empirically verifiable).
  - Argues that the distinction is unclear and untenable.
- **Verification Theory of Meaning:**
  - Questions the idea that meaningful statements are those that can be empirically verified.
  - Advocates for a more holistic view of language and meaning.
- **Holism:**
  - Introduces the idea of holism, suggesting that the meaning of a sentence is derived from its place in a broader web of interconnected beliefs.
  - Rejects the isolation of individual statements for analysis.
- **Indeterminacy of Translation:**
  - Discusses the indeterminacy of translation, proposing that there can be multiple ways to translate one language into another without a unique correct translation.
- **Empirical Content and Reductionism:**
  - Criticizes the reductionist attempt to reduce all statements to observational statements with empirical content.
  - Asserts that observational statements alone cannot determine the truth or falsity of scientific theories.
- **Impact on Philosophy:**

- Quine's ideas have had a profound impact on the philosophy of language and epistemology.
- The rejection of the analytic-synthetic distinction and the emphasis on holism continue to shape contemporary discussions in philosophy.

Criticizes the paradigm of logical positivism while being empiricist about it

Rudolf Carnap - The Elimination of Metaphysics Through Logical Analysis of Language

#Philosophy

Rudolf Carnap: a prominent philosopher in the early-to-mid 20th century, associated with the Vienna Circle and logical positivism. In his work, he aimed to address the issues he saw with traditional metaphysics by advocating for the elimination or reduction of metaphysical problems through logical analysis of language.

The key idea behind Carnap's position is that many metaphysical problems arise due to the misuse or misunderstanding of language. He argued that by clarifying and analyzing the language we use to express philosophical concepts, we can dissolve or eliminate apparent metaphysical problems. Carnap believed that these problems often arise from the use of unclear or ambiguous language, leading to confusion and pseudo-problems.

Carnap's approach can be understood through the following steps:

- **Language Analysis:** Carnap emphasized the importance of precise and clear language.
- **Logical Syntax:** Carnap was interested in the logical structure of language. He introduced the idea of "logical syntax" as a means of studying the formal structure of languages. By doing so, he sought to reveal the logical relationships between different statements and to establish a clear and unambiguous framework for expressing meaningful propositions.
- **Verification Principle:** Carnap, along with other logical positivists, advocated the verification principle as a criterion for meaningfulness. According to this principle, a statement or proposition is meaningful only if it can be empirically verified or is logically necessary. Anything that cannot be verified through empirical observation or logical analysis was considered meaningless, including many traditional metaphysical claims.
- **Empirical Content:** Carnap believed that genuine knowledge should be grounded in empirical content, and he was critical of speculative metaphysical

claims that lacked empirical support. By emphasizing empirical content, he aimed to redirect philosophy toward the analysis of language and the clarification of concepts that have a clear connection to observable phenomena.

Introduction

> "In the strict sense, however, a sequence of words is *meaningless* if it does not, within a specified language, constitute a statement. It may happen that such a sequence of words looks like a statement at first glance; in that case we call it a *pseudo-statement*"

There a two kinds of pseudo-statements:

1. Either they contain a word which is erroneously believed to have meaning
2. Or the constituent words are meaningful, yet they are put together in a counter-syntactical way, so that they don't yield a meaningful statement

The Significance of a Word

> "A word which (within a definite language) has a meaning, is usually also said to designate a concept; if it only seems to have a meaning while it really does not, we speak of a "pseudo-concept.""

Carnap argues that words can lose their original meaning over time without gaining a new one and thus become "pseudo-concepts"

The meaning of a word is determined by its "criterion of application"

- By the relations of deducibility entered into by its elementary sentence-form, by its truth-conditions, by the method of verification
- The stipulation of the criterion takes away one's freedom to decide what one wishes to "mean" by the word
- If no criterion of application is stipulated, then nothing is asserted by the sentences in which it occurs, they are but pseudo-statements

Carnap puts it like this:

Let "a" be any word and "S(a)" the elementary sentence in which it occurs. Then the sufficient and necessary condition for "a" being meaningful may be given by each of the following formulations, which ultimately say the same thing:

1. The *empirical criteria* for "a" are known
2. It has been stipulated from what protocol sentences "S(a)" is *deducible*
3. The *truth conditions* for "S(a)" are fixed
4. The method of *verification* of "S(a)" is known

Metaphysical Words without Meaning

Carnap posits that the word and "metaphysical term": "principle", does not meet the aforementioned requirement thus making it "meaningless"

- Metaphysicians fail to stipulate a precise criterion of application to the word "principle" and since they do not use it to mean what the original word "principium" meant it has become "an empty shell"

Carnap also makes the case for "God" not meeting his requirements in certain cases (allows the *mythological* use though)

- The *theological* usage of the word falls between its mythological and metaphysical usage, but predominantly still becomes "meaningless"

Carnap gives a bunch more superficial examples like "the Idea", "the Absolute", "spirit" etc., which he believes to be metaphysical and therefore devoid of meaning

Significance of A Sentence

There is a second kind of pseudo-statement, one which consists of meaningful words, but which are put together in such a way that nevertheless no meaning results.

Example:

1. Caesar is and
2. Caesar is a prime number

1 is just syntactically incorrect
2 is syntactically correct but is meaningless because "prime number" is a predicate of numbers not of a person

Carnap argues that from a logical points of view the grammatical syntax of natural language is inadequate, precisely because it is possible to construct such nonsense sentence while still being syntactically correct

> "In a correctly constructed language, therefore, all nonsensical sequences of words would be of the kind of example (1)"

Carnap then says that if the statements of metaphysics are pseudo-statements then metaphysics could not even be expressed in a logically constructed language.

Metaphysical Pseudo-Statements

Carnap illustrates some logical construction of sentences and again disses metaphysics. He says later on: "We do not regard metaphysics as "mere speculation" or "fairy tales". The statements of a fairy tale do not conflict with logic, but only with experience; they're perfectly meaningful, although false."

- Basically says that metaphysical statements aren't even false, they're impossible or "meaningless"

Repeats that statements have to be verifiable to become meaningful

Meaninglessness of all Metaphysics

Subtitle speaks for itself. Carnap is not a fan

Meaningful statements are divided into the following kinds:

- Statements which are true solely by virtue of their form ("tautologies" according to Wittgenstein; correspond approximately to Kant's "analytic statements")
  - Logic, mathematics
- Negotiations of such statements ("contradictions")
  - Self contradictory = false by virtue of their form

Metaphysics as Expression of an Attitude toward Life

Starts speaking about music, art and science and how metaphysics apparently confuses these and ends up doing nothing

Carnap does not hate Nietzsche's metaphysics apparently:

> "Our conjecture that metaphysics is a substitute, albeit an inadequate one, for art, seems to be further confirmed by the fact that the metaphysician, who perhaps had artistic talent to the highest degree, viz. Nietzsche, almost entirely avoided the error of that content"

Carnap - Psychology in Physical Language-DESKTOP-NQIJK4K

#Philosophy

Introduction. Physical Language and Protocol Language

Opens quite strongly:

> "IN WHAT FOLLOWS, we intend to explain and to establish the thesis that every sentence of psychology may be formulated in Physical language. To express this in the material mode of speech: all sentences of psychology describe physical occurrences, namely, the physical behavior of humans and other animals."

and

> "This is a sub-thesis of the general thesis of *physicalism* to the effect that physical language is a *universal language*, that is, a language into which every sentence may be translated"

If our thesis is correct, the generalized sentences of psychology, the laws of psychology, are also translatable into physical language. They are thus physical laws.

The Forms of Psychological Sentences

- A singular psychological sentence e.g. “Mr. A was angry at noon yesterday” (an analogue of the physical sentence, “Yesterday at noon the temperature of the air in Vienna was 28 degrees centigrade”), is concerned with a particular person at a particular time
- General psychological sentences have various forms, of which the following two are perhaps the most important
    1. A sentence may describe a specific quality of a specific kind of event, e.g. “An experience of surprise always (or: always for Mr. A, or: always for people of such and such a society) has such and such a structure.” A physical analogy would be: “Chalk (or: chalk of such and such a sort) always is white.”
    2. universal-conditional statements concerning sequences of events, that is, of causal laws. For instance,
        1. “When, under such and such circumstances, images of such and such a sort occur to a person (or: to Mr. A, or: to anyone of such and such a society), an emotion of such and such a sort always (or: frequently, or: sometimes) is aroused.” A physical analogy would be: “When a solid body is heated, it usually expands.”

Phenomenology claims to be able to establish universal synthetic sentences which have not been obtained through induction. These sentences about psychological qualities are, allegedly, known either *a priori* or on the basis of some single illustrative case

- In our view, knowledge cannot be gained by such means. We need not, however, enter upon a discussion of this issue here, since even on the view of phenomenology itself, these sentences do not belong to the domain of psychology.

In physics it sometimes seems to be the case that a general law is established on the basis of some single event. For instance,

- if a physicist can determine a certain physical constant, say, the heat-conductivity of a sample of some pure metal, in a single experiment, he will be convinced that, on other occasions, not only the sample examined but any similar sample of the same substance will, very probably, be characterizable by the same constant.
- But here too induction is applied. As a result of many previous observations the physicist is in possession of a universal sentence of a higher order which enables him in this case to follow an abbreviated method. This higher-order sentence

reads roughly: “All (or: the following) physical constants of metals vary only slightly in time and from sample to sample.”

The situation is analogous for certain conclusions drawn in psychology. If a psychologist has, as a result of some single experiment, determined that the simultaneous sounding of two specific notes is experienced as a dissonance by some specific person A, he infers (under favorable circumstances) the truth of the general sentence which states that the same experiment with A will, at other times, have the same result

- Here too the inference from a singular sentence to a general one is only apparent. Actually, a sentence inductively obtained from many observations is brought into service here

It thus remains the case that every general sentence is inductively established on the basis of a number of singular ones.

Finally, we must consider sentences about psycho-physical interrelations, such as for instance, the connection between physical stimulus and perception. These are likewise arrived at through induction, in this case through induction in part from physical and in part from psychological singular sentences. **The most important sentences of gestalt psychology belong also to this kind.**

General sentences have the character of hypotheses in relation to concrete sentences, that is, the testing of a general sentence consists in testing the concrete sentences which are deducible from it. A general sentence has content insofar and only insofar as the concrete sentences deducible from it have content. Logical analysis must therefore primarily be directed towards the examination of the latter sort of sentences

Sentences About Other Minds

The epistemological character of a singular sentence about other minds will now be clarified by means of an analogy with a sentence about a physical property, defined as a disposition to behave (or respond) in a specific manner under specific circumstances (or stimuli).

A Sentence about a property of a physical substance

Example: I assert the sentence P_1: “This wooden support is very firm.”

A Sentence about a condition of some other mind.

Example: I assert the sentence P_1: "Mr. A is now excited.”

These are two different ways in which sentence P_1 may be derived. We shall designate them as the “ rational ” and the “ intuitive ” methods

The rational method consists of inferring P_1 from some protocol sentence p_1 (or from several like it), more specifically, from a perception-sentence

- about the shape and color of the wooden support.
- about the behavior of Mr. A, e.g. about his facial expressions, his gestures, etc., or about physical effects of A’s behavior, e.g. about characteristics of his handwriting

In order to justify the conclusion, a major premise O is still required, namely the general sentence which asserts that

- when I perceive a wooden support to be of this color and form, it (usually) turns out to be firm. (A sentence about the perceptual signs of firmness.)
- when I perceive a person to have this facial expression and handwriting he (usually) turns out to be excited. (A sentence about the expressional or graphological signs of excitement.)

Lau says:

- $P_1 \Leftrightarrow (p_1 \wedge O)$
- "P1 is true if and only if both p1 and O are true simultaneously." It's expressing an equivalence relationship between P1 and the conjunction of p1 and O.
    - Creds to ChatGPT

The content of P_1 does not coincide with that of p_1, but goes beyond it. This is evident from the fact that to infer P_1 from p_1 O is required. The cited relationship between P_1 and p_1 may also be seen in the fact that under certain circumstances, the inference from p_1 to P_1 may go astray. It may happen that, though p_1 occurs in a protocol, I am obliged, on the grounds of further protocols, to retract the established system sentence P_1. I would then say something like, “I made a mistake. The test has shown:

- that the support was not firm, even though it had such and such a form and color.”
- that A was not excited, even though his face had such and such an expression.”

In practical matters the intuitive method is applied more frequently than this rational one, which presupposes theoretical knowledge and requires reflection.
In accordance with the intuitive method, P_1 is obtained without the mediation of any other sentence from the identically sounding protocol sentence p_2.

- “The support is firm.”
- “A is excited.”

Consequently, one speaks in this case of *immediate perceptions*

- of properties of substances, e.g., of the firmness of supports.
- of other minds, e.g., of the excitement of A

Here too we can best clarify the difference by considering the possibility of error. It may happen that, though p_2 occurs in my protocol, I am obliged, on the basis of further protocols, to retract the established system sentence P_1. I would then say "I made a mistake. Further tests have shown:

- that the support was not firm, although I had the intuitive impression that it was."
- that A was not excited, although I had the intuitive impression that he was."

Bunch of text wont write, so picture:

respective fields, the majority of them nowadays would give us thoroughly non-analogous answers. The identity of the content of P_2

and of the content of the physical sentence P_1 would be agreed to as a matter of course by all physicists.

and of the content of the psychological sentence P_1 would be denied by almost all psychologists (the exceptions being the radical behaviorists).

The contrary view which is most frequently advocated by psychologists is that, "A sentence of the form of P_1 asserts the existence of a state of affairs not identical with the corresponding physical structure, but rather, only accompanied by it, or expressed by it. In our example:

P_1 states that the support not only has the physical structure described by P_2, but that, besides, there exists in it a certain force, namely its *firmness*.

This firmness is not identical with the physical structure, but stands in some parallel relation to it in such a manner that the firmness exists when and only when a physical structure of the characterized sort exists.

Because of this parallelism one may consider the described reaction to certain stimuli—which is causally dependent upon that structure—to be an *expression* of firmness.
Firmness is thus an occult property, an obscure power which stands behind physical structure, appears in it, but itself remains unknowable."

P_1 states that Mr. A not only has a body whose physical structure (at the time in question) is described by P_2, but that—since he is a *psychophysical being*—he has, besides, a consciousness, a certain power or entity, in which that excitement is to be found.
This excitement cannot, consequently, be identical with the cited structure of the body, but stands in some parallel relation (or in some relation of interaction) to it in such a manner that the excitement exists when and only when (or at least, frequently when) a physical, bodily structure of the characterized sort exists.
Because of this parallelism one may consider the described reaction to certain stimuli to be an *expression* of excitement.

Excitement, or the consciousness of which it is an attribute, is thus an occult property, an obscure power which stands behind physical structure, appears in it, but itself remains unknowable."

P_2: a dispositional statement (logical behaviorism)

This view falls into the error of a hypostatization[1] as a result of which a remarkable duplication occurs: besides or behind a state of affairs whose existence is empirically

determinable, another, parallel entity is assumed, whose existence is not determinable. (Note that we are here concerned with a sentence about other minds.) But—one may now object—is there not really at least one possibility of testing this claim, namely, by means of the protocol sentence p2 about the intuitive impression of the firmness of the support?

The objector will point out that this sentence, after all, occurs in the protocol along with the perception sentence p1. May not then a system sentence whose content goes beyond that of P2 be founded on p2? This may be answered as follows.

- A sentence says no more than what is testable about it. If, now, the testing of Pl consisted in the deduction of the protocol sentence p2, these two sentences would have the same content. But we have already seen that this is impossible

There is no other possibility of testing Pl except by means of protocol sentences like p1 of p2. If, now, the content of Pl goes beyond that of P2, the component not shared by the two sentences is not testable, and is therefore meaningless. If one rejects the interpretation of Pl in terms of P2, Pl becomes a metaphysical pseudo-sentence.

---

1. Wtf↩

Bechtel - Philosophy of Science

Introduction: The Origins of Logical Positivism

The logical positivists were not attempting to account for all scientific activity

1. They distinguished between *the context of discovery* in which scientific hypotheses were developed and the *context of justification* in which they were rationally assessed

Even in the context of justification the Positivists recognized that practicing scientists often do not adhere to the canons of formal logic. What the Logical positivists maintained was that justificatory reasoning of scientists, if it was good, could be *reconstructed* to accord with the construal of scientific reasoning based on modern logic

- They were proposing normative standards for science

The Verifiability Theory of Meaning

The logical positivists attributed many of the confusions of science, particularly in social and behavioral sciences to unclarity of language

In their discussions of meaning the positivists followed the classical empiricists in linking knowledge to experience, but they advocated one important change.

- The classical Empiricists treated ideas as the units of thinking and view theses ideas as causal products of sensory experience
- The Logical Positivists rejected ideas as fuzzy entities. Rather they took linguistic entities --sentences and words-- to be the basic vehicles of meaning

They proposed the criterion of verification to explain how these linguistic entities could be appropriately related to experience

- *the meaning of a sentence was the set of conditions that would show that the sentence was true*

Because only sentences and not individual words could be true or false, the meaning of words had to be analyzed in terms of their roles in sentences

- This account of meaning became known as the **verifiability theory of meaning**

Some sentences according to LP could be directly verified through experience, these were referred to as *protocol sentences* or *observation sentences*

- Of course they disagreed quite extensively among themselves on which sentences counted as such
- Carnap restricted observation sentences to those characterizing our phenomenal experience (e.g. "I am sensing a blue color patch now")

Other sentences cannot be verified directly through experience (particularly sentences containing theoretical terms e.g. force)

- To explicate the meaning of these terms the positivists focused on ways in which the truth or falsity of sentences using these terms could be determined indirectly via other sentences that were observational
- Thus translation consists of biconditional sentences that assert that one statement (the theoretical statement) is true if and only if another, possibly complex statement (the observation statement) is true