Lecture notes for Philosophy of Cognitive Science

An upper level under-graduate course taught by:
Kelly Alexandra Roe

2010

Introduction

In 2010 I was subcontracted to teach (lecture, tutor, grade) around $\frac{3}{4}$ of a course in Philosophy and Cognitive Science at Macquarie University, Sydney.

The tenured Professor had received an Australian Research Council grant to research psychopathy so she subcontracted most of of the teaching tutoring and grading work to me while I waited for the ANU to get my research out for external examination.

I wrote lecture note outlines to accompany the following text-book (which had already been ordered):

Clarke, A. (2002). Mindware: An introduction to the philosophy of cognitive science. Oxford University Press.

I have adapted them from power-point presentations that were uploaded to the online learning platform, and distributed as handouts in class.

The notes are basic in outline. They served as a prompt for me and a reminder to students of the reading they had done in advance. They were distributed in class and by way of online learning platform.

I was told by the department chair not to assume that students had taken previous courses in psychology, philosophy of psychology, or philosophy of mind.

Thanks to students for attending, engaging in such high-quality discussion in both the lectures and tutorials, and writing such wonderful essays in response to the content of this course.

I believe that Macquarie retains the audio recordings of my presentation of the lectures.

Contents

1. Introduction

What is cognitive science?

- Thinking, reasoning, inferring, choosing, deciding, willing, intending, loving, fearing, hoping, wishing, imagining, seeing, hearing, smelling, tasting, feeling, experiencing, dreaming...
- Inter-disciplinary science of cognition
  - Philosophy
  - Cognitive psychology
  - Computer science (artificial intelligence and robotics)
  - Linguistics
  - Anthropology
  - Neuroscience
  - Researchers and theorists in related fields such as education, developmental psychology, ethology etc.

What is the role of philosophy in cognitive science?

- This is controversial. Two main views:
  1. There is a radical divide between philosophy and the natural sciences
  2. There is a continuum between philosophy and the natural sciences

1. The radical divide

- Some theorists maintain that there is a radical divide between philosophy and the natural sciences
  - Soul vs matter
  - Mind vs matter
  - Value vs matter
  - Meaning vs matter
- The findings of science are not relevant to answering philosophical questions
- Science misses the point / changes the subject
  - If this is so it may be that it is because philosophical theories aren't testable

2. The continuum

- There is some kind of reciprocal relationship or continuity between philosophy and the natural sciences
- Look to science for data that is relevant for answering philosophical questions
  - If this is so science may be data collection for the philosophical theorist
- Look to philosophy for questions, hypotheses, and / or theories
  - If this is so there may be no more philosophy with scientific progress
- In doing science we have to start somewhere
- Better to start from some place plausible rather than from some place implausible
- So let us start with some 'common-sense' intuitions
  - I mean to say I will try and articulate some of the following, hopefully in a way that seems intuitive to you

Mental states as propositional attitudes

- Each of these mental states seems to be ABOUT something (p):

  > Thinking that p, reasoning that p, inferring that p, choosing p, deciding p, willing p, intending p, loving p, fearing that p, hoping that p, wishing that p, imagining p, seeing p, hearing p, smelling p, tasting p, experiencing p, dreaming p...

- What they are about (p) is the CONTENT of the mental state
- Mental contents are thought to be PROPOSITIONAL ATTITUDES
- Propositions are (for our purposes) abstract meanings or informational contents
  - 'The sun is hot'
  - 'Hot, the sun is'
  - Sentences in other language that are synonymous
- Different sentences, phrases, expressions, thoughts that have the same meaning, informational, or propositional content

Some features of the propositional view

- One person can entertain the very same propositional content at different points in time
- Different people can entertain the very same propositional content
- Speakers of different languages can entertain the same propositional content (say the same things or think the same thoughts)

Folk psychology (aka belief-desire psychology or 'common-sense psychology')

- We can make fairly good predictions about what people will do by appealing to what they believe (the way they represent the world to be) and what they desire (the way they are motivated to alter the world)
- We also explain actions by appealing to what people believed (represented) and what people desired (were motivated to do)

- It seems common-sense that (roughly) ‘an agent will act so as to satisfy or obtain their strongest desire under the assumption that their beliefs are true’
- No other (scientific) theory seems to capture the relevant predictions

Questions

- What enables us to attribute mental states to others / employ folk-psychology?
- What does this ability show us about the structure of the mind?
- What is the status of folk-psychology compared to scientific theories in physics, chemistry, biology, cognitive psychology etc?
- What does folk-psychology show us about the nature of mental states?
- Are (all?) mental states usefully thought of as propositional attitudes?

2. Dualism

Reminder of the cognitive phenomenon

- Thinking, reasoning, inferring, choosing, deciding, willing, intending, loving, fearing, hoping, wishing, imagining, seeing, hearing, smelling, tasting, feeling, experiencing, dreaming...

Preview

- Theories of the mind-body (mind vs matter) relationship that we will be looking at:

  1. Dualism
  2. Behaviourism
  3. Identity theory
  4. Eliminativism
  5. Functionalism

Dualism (mind is distinct from matter)

- *Why* is dualism thought to be plausible?
  - Introspection. It just seems (upon introspection) that the mind (or perhaps the soul) is different or distinct from matter
  - How could... How could certain phenomenon (the way that my pain feels or inferential reasoning, for example) arise from 'matter nicely orchestrated'?
- There are two main kinds of dualism

1. Substance dualism
2. Property dualism

1. Substance dualism

- The mind (soul, spirit) is a substance (kind of stuff, object, or thing) that is fundamentally or radically different from matter

2. Property dualism

- Mental states are non-physical (immaterial) properties (states) of the physical brain or body

Objects vs Properties

- In order to understand the difference between substance and property dualism we will need to take a look at the difference between substance and property
- Water is thought to be a substance
- Liquidity is thought to be a physical property of water (and substances other than water)
- Mental states (according to the dualist) are different from water or from fluidity in being non-physical or immaterial

Dualism and causal interaction

- There are three main views on how the mental (substance or property) causally interacts with or relates to material or physical substance or property
  1. Interactionism
     - Two-way causal interaction between body and mind (e.g., Descartes)
  2. Epi-phenomenalism

  - Matter causes changes to matter but mind is itself causally impotent (the shadow analogy)

3. Paralellism
   - No causal interaction between body and mind (e.g., Leibniz)

Problems for dualism

- Providing an account of non-material stuff / properties
- Understanding how causal interaction is possible (on the interactionist views)
- A simpler explanation posits only one kind of stuff (or property) rather than two (Morgan's cannon, Ockham's razor)
- Analogy with mind and vital spirit as science progresses and commonsense is revised

Currently

- No many defenders of substance dualism
- A few defenders of property dualism for consciousness (more on that later) e.g., Block, Chalmers
- Most would say that dualism has been abandoned as the result of scientific advances (taking the vitalism analogy very seriously indeed)
- We will return to this when we look at consciousness

3. Behaviourism

Behaviourism (mental states are behavioural dispositions)

- Two main kinds:
    1. Analytical behaviourism
    2. Methodological behaviourism

Analytical behaviourism

- Primarily a thesis about how we should analyze mental discourse
- E.g., to say ‘x is in pain just is to say that ‘if x’ were placed in these circumstances then x would be disposed to...’
- Since mental language refers to dispositions to behave, mental states just are behaviours or dispositions to behave

Methodological behaviourism

- Different psychoanalytic / psychodynamic theories of the mind seemed ‘unscientific’
- In order to become a ‘real science’ ‘just like physics’ the best bet for psychology is to become the science of behaviour (e.g., Skinner, Watson)
- Strictly speaking, methodological behaviourists don’t need to be analytical behaviourists

- In practice most justify their view by appealing to analytical behaviourism, however

Why is behaviourism thought to be plausible?

- Mostly because of LEARNING
- We don't observe others mental states directly - so how do we learn to label our own?
- The importance of behaviour as a source of evidence for figuring out what mental state a person (including ourself) is in

Problems for analytical behaviourism

- Providing a fuller account of the dispositional profiles
    - Problems cashing out the dispositional profile (concern it might be infinite or circular)
    - The thought that the disposition might best be thought of as an inner state of the brain
- Doesn't seem so plausible as an account of the felt quality of experience (qualia, p-consciousness)

Problems for methodological behaviourism

- The birth of cognitive psychology and the cognitive revolution showed that psychology didn't have to be about behaviours / behavioural dispositions in order to be scientific!

Currently...

1. Within psychology it is often thought that cognitive psychology *replaced* behaviourism as a methodological paradigm
2. Functionalism may be viewed as an *extension or* development of analytical behaviourism, however (as we shall see)

3. In practice methodological behaviourism is alive and well in certain areas (e.g., animal behaviour in ethology, behavioural change especially in children, and the intellectually handicapped in clinical psychology)

4. Identity Theory

Identity theory (mental states are brain states)

Identity theory aka: The ‘Australian Thesis’

- Two main kinds:
    1. Type-type
    2. Token-token (arose later - a strategic retreat)

Why is the identity theory thought to be plausible?

- Just as science showed us that lightening just is a certain kind of electrical discharge...
- Science is showing us that mental states just are states of the brain
- Neuro-science is appealing to many. It seems to ‘pop the hood’ on behavioural dispositions
- We see lots of pictures of the neural correlates of various mental activities in journals

Type-type identity theory

- Types of mental states are types of brain states (e.g., pain = C fibres firing)

Problems with type-type identity theory

- If mental states = brain states then beings without brains (e.g., computers, robots, certain kinds of aliens, angels, god) can't have mental states
- If dolphins don't have C-fibres (or whatever brain state we are in when we are in pain) then they can't be in pain and this seems counter-intuitive to most
- Leibniz law objections
  - If you have x and y and want to know whether x=y (where = is the identity relation that each object bears only to itself then if you can find a property that x has that y lacks (or vice versa) then you can conclude that x does not = y. Some candidates:
    - My pain is in my toe but my brain is in my head
    - My beliefs can be true or false but my brain states can't be
  - Responses
    - It might *seem to you* that they have different properties
    - But that is a feature of *you* and you are *wrong or misguided*
    - E.g., Your brain *represents* bodily damage in your toe and that state of your brain just is pain. So the pain is in your brain but your brain *represents* the pain as being in your toe
    - Similarly, what your brain *represents* can be true or false (e.g., referred pain) and beliefs are just your brain *representing* things to be a certain way
  - Currently a number of neuroscientists / cognitive neuroscientists think that type-type identities have been or will be made between mental or cognitive states and neuro-physiological states
  - A number of other neuroscientists or cognitive neuroscientists think that actually type-type identities haven't been as forthcoming as we would have hoped
  - The later has fuelled the two further developments that we will look at - eliminativism and functionalism.

5. Eliminativism

Eliminativism (there aren't any mental states)

- Folk-psychology and our common-sense understanding of mental states involves our committing to a certain view of their nature (e.g., that mental states are types of brain states)
- It turns out that (according to some neuroscientists) mental states aren't correlated with brain states
- Therefore, neuroscience has shown us that there aren't any mental states. Just like science showed us that there isn't any phlogiston
- Paul and Patricia Churchland think that as neuroscience matures the vocabulary of neuroscience will come to replace the vocabulary of folk-psychology / mentalistic discourse
- Neuroscience textbooks don't talk about (have a chapter on) 'belief' so it seems that some of our mentalistic discourse already has been eliminated from neuroscience, at least.

Problems for eliminativism

- We can't just eliminate mentalistic discourse from our everyday lives and go on with business as usual.
- Neuroscience seems to be at the wrong grain to capture the predictions that can be made from folk-psychology (e.g., that people will turn up to class this week because of certain beliefs and desires they have. We will return to this.
- Perhaps any theory that commits us to concluding that 'there aren't any mental states' must be false

Problems for mentalistic discourse

- But then isn't mentalistic discourse scientific if it isn't open to being falsified?
- What does that imply for a science of cognition?

6. Functionalism

Functionalism (mental states are functional role states)

Preview

- Functionalism, functions, functional roles
- Machine tables as functional role characterisations
- Semantics, syntax, reducing semantics to syntax
- Kinds of functionalism
- A concern for machine functionalism
- There has been a pendulum between dualism and materialism through history
  - Functionalism is an attempts to avoid the pendulum
- Wouldn't it be nice if there were a theory to capture what is plausible in what went before while avoiding some of the problems?
  - Functionalism is an attempt to do so

What are functions?

- Think of mathematical functions:
  - -, +, =, x (mathematical operators)
- Or logical functions:
  - $\neg, \vee, \rightarrow$, (logical operators)

- Transformations:
    - (Modus ponens, or disjunction introduction)
- Or syntactic functions:
    - Rules of combining words into sentences
    - Transformations (e.g., present to past tense)

What is a functional role?

- In order for x to count as a state of belief x must play the functional role of belief
- Functional roles are abstract, structural, formal, or syntactical properties

A Machine Table

- Example of a soda machine that takes 50c and 1$ coins. Soda costs $1.50 and the machine will give change
- 4 states of the machine:
    - State 0
        * If $1 is input then goto state 1
        * If 50c is input then goto state 2
    - State 1
        * If 50c is input then output soda and goto state 0
        * If $1 is input then output soda and output 50c and goto state 0
    - State 2
        * If $1 is input then output soda and goto state 0
        * If 50c is input then goto state 3
    - State 3
        * If $1 is input then output soda and output 50c and goto state 0

    - ∗ If 50c is input then output soda and goto state 0

Functionalism and machine tables

- The machine table specifies different (internal) states of the coke machine (0, 1, 2, and 3)
- Each state is defined by its abstract structural, formal, or syntactic relation to:
    1. Inputs ($1 and 50c)
    2. Internal states (goto)
    3. Outputs (soda, change)
- Functionalism is thus a tripartite, or three-part theory

Semantics vs syntax

- A semantics for a language is the meaning, or informational content that the syntax provides rules for manipulating
- A semantics for logic would replace content-less variables (p, q etc) with semantic contents (e.g., Socrates, man)
- Thus we have a distinction between content / meaning and rules that govern content / meaning transitions
- One (controversial) thought is that ‘if you take care of the syntax then the semantics will take care of itself’
- The thought is that semantic content (e.g., ‘dog’) can be characterised syntactically with respect to:
    - Typical input (dogs)
    - Inferences it licenses (is not a cat), the relation it bears to other states (e.g., desires)
    - The output that is produced (e.g., petting)
- We will return to look at machine intelligence
    - Can programming an appropriate syntax give machines content to think about (genuine understanding)?

Kinds of functionalism.

- How do we specify the functional role of the different kinds of mental states?
  - Machine functionalism - look to logic / syntactic transformation rules
  - Analytic functionalism - look to common-sense folk-psychology
  - Empirical functionalism - look to science (e.g., cognitive psychology, biological psychology)
  -

A concern

- 'Chauvinism' was an objection to the type-type identity theory
  - Beings with different brains or no brains could have mental states
- 'Excessive liberalism' is an objection to machine functionalism
  - A bucket of river water warming in the sun can probably be described as instantiating any computational description that you care to think of
- Getting the balance between these is tricky

7. Mind as meatware

Mind as wetware

Preview

- The role / realizer distinction
- Multiple realizability
- Software / hardware
- Mindware / wetware

Roles vs realizers

- We have seen that functionalists think that mental states are functional roles
- The functional role that is thought to be relevant depends on the version of functionalism (whether the functional role is to be given by common-sense, computational specification, or by the empirical biological sciences)
- It is only in virtue of the state playing the relevant functional role that the state would count as being a mental state
- Functionalists identify mental states with the functional role being filled rather than with whatever it is that happens to fill the functional role
- That is what makes functionalism different from the view that mental states are to be identified with (are one and the same as) whatever it is that happens to realize the role (e.g., that they are brain states or immaterial states)

- Example: Consider a doorstop. A doorstop (let us suppose) is whatever it is that plays the doorstop role.
- A shoe, bag, block of wood, rolled up newspaper etc isn't a doorstop - except insofar as it is realizing or instantiating the doorstop role. That is to say it is being used as a doorstop.
- If you were to go a step further and say that that particular shoe really is a doorstop (even when it is not being used as a doorstop) then that would be token-token identity theory. That means to say this particular (token of) a shoe is (an 'is' of identity) a particular instance or token of a doorstop.
- If you were to go a step further and say that shoes are doorstops that would be type-type identity theory regarding the relationship between shoes and doorstops

Multiple realizability

- While the role of the states can be specified by their inputs, internal relations to each other, and their outputs the realizers (particular things) that fill or instantiate the role can be made of glass, copper, tin, plastic, immaterial souls or ghosts etc
- Thus functionalism (strictly speaking) avoids the pendulum swing between our having to choose materialism or dualism by remaining neutral or agnostic as to the nature of the realizers
- The realizers could be neural states or silicon states of a computer or nitrogen hydraulics of aliens or immaterial states of a Cartesian soul or animus from the breath of gods
- Mental states are thus multiply realizable
- Which means that we can have a science of the mind / cognition without worrying about neurons or the nature of the hardware.

Mind is to brain as software is to hardware

- The intuitive idea is that the same software programme (e.g., Microsoft word) can run on different hardware (e.g., PC, mac)

- Though they acknowledge that hardware constrains software (e.g., you can't run Microsoft word on water)
- We can consider features of the Microsoft word programme abstractly enough so that the different hardware is irrelevant
  - Critics maintain that differences in hardware make important differences to relevant features of the software (e.g., processing time)

Mindware / wetware

- Similarly, while some cognitive psychologists maintain that the mind program can be studied in abstraction from the neural implementation...
- They allow that the neural basis does impose some constraints (e.g., on processing time) but they don't think that those constraints are particularly relevant for understanding the mind
  - Critics maintain that neurological differences will turn out to be crucially important and they cannot be ignored

Multiple realizability and types

- So while multiple realizability is typically thought to be a feature of functionalism, critics maintain that it doesn't hold up. It is hard to know how the science will go.

Types of states

- There is an issue around what types of mental states there are
  - E.g., folk psychology considers 'memory' a type of mental state but folk psychology considers 'iconic visual sensory register'. 'semantic memory'. 'episodic memory,' etc to be different types of mental states
- There is an issue around what types of brain states there are
  - E.g., activation, or a more particular kind of activation?
- Maybe if we got both of those right there would be type-type correlations

8. Consciousness

Consciousness

- What is consciousness? Some candidates:
  - Awakeness
  - Self awareness
  - Availability for verbal report
  - Availability for the control of intentional action
  - Qualia (qualitative experience, phenomenal awareness)
- Used to be regarded as 'off limits' for science or scientific research
- Now if you read the cognitive neuroscience literature you might well think that the problem is solved
- Fashionable topic, currently
- Some people think that much of the science misses the point
- Other people think that the philosophical notion of consciousness needs to be rehabilitated else eliminated
- Distinction between A (access) consciousness and P (phenomenal) consciousness
- The (comparatively) 'easy' and 'hard' problems of consciousness

Awakeness

- The distinction or difference between being asleep and being awake
- Relevant for anesthesiology

- 'She's unconscious' and 'she's asleep' seem to be used synonymously or treated as synonyms
- We are learning much about the (relatively) primitive brain structures that regulate sleep and wakefulness
- *But it seems that we can have conscious experiences while asleep - e.g., the experiences we have while dreaming.*

Self awareness

- The capacity for 'meta-cognition' - to reflect on our cognitive states and experiences
- To have a sense of ourself as persisting through time, having different projects and preferences and dreams
- People with dissociative identity disorder (formerly known as multiple personality disorder) experience a 'fragmented sense of personal identity' (multiple selves). This is thought to be a disorder of consciousness
- *Small children and animals lack a sense of self-awareness but still have experiences*

Availability for verbal report

- Often the best way to know what a person is experiencing is to ask them
- We seem to be able to report on our conscious experiences
- *Seems to suggest that those who lack verbal capacity (e.g., animals and small children) lack conscious experience insofar as their experiences are not available for verbal report*
- *Seems possible to promptly FORGET a conscious experience. E.g., would you rather take a drug for surgery that blocked the experience of pain or blocked the availability of the experience to verbal report?*

Availability for control of intentional action

- People can often use their conscious experiences to guide a diversity of plans, projects, and goals

- E.g., psychological experiments where people press a key when they have certain experiences. These experiments are interpreted as showing us whether people have consciously experienced the stimuli
- It is unclear how much animals and small children are able to control intentional action yet it seems intuitive to most of us that they are conscious
- Would you rather lack the experience of pain or lack the availability of the experience of pain to guide your action?

Qualitative experiences or qualia (singular - 'quale')

- Probably impossible to define
- Can gesture towards the phenomena... And hope that people get the intuition
    - Descartes - cognito ergo sum
    - The particular way that your experiences feel or seem to you from your first person point of view
    - The sum total of your experience right now is your 'phenomenal field'

Some scientific research on consciousness

- The binding problem (the unity of the phenomenal field)
- Blindsight
- Dorsal vs ventral processing

Binding

- The experiential field seems to be unified
    - Within modalities. E.g., objects near and far away seem to be present at the same time
    - Between modalities. E.g., vision and audition

- The binding problem has to do with how the brain manages to unify sensory modalities to underwrite the unity of experience
  - E.g., the 40hz thesis (when different regions fire in this frequency the information presents as bound in the phenomenal field)

Blindsight

- Cortical damage results in a schomata (blind spot) in the visual field
- Can present information to that region
- Patients report no conscious visual experience
- But when they are forced to 'guess' their guesses are above chance

Dorsal vs ventral processing

- Dorsal stream - 'where' (spatial location) - guidance of action
- Ventral stream - 'what' (identification / recognition) - perception involved in visual awareness
- So action without awareness (ventral deficit) or awareness without appropriate action (dorsal deficits)

Ned Block

- A consciousness (access consciousness) - information poised to control action and verbal report
- P consciousness (phenomenal consciousness) - qualia (qualitative experience or the felt quality of experience)

David Chalmers

- The (comparatively) 'easy problem' of consciousness - learning about states that are poised to control action and verbal report
- The (comparatively) 'hard problem' of consciousness - learning about qualia

- ‘What is striking is that it is only that final target (qualia) that threatens to present any special kind of problem for our standard modes of cognitive scientific explanation and understanding’ (Textbook)
-

Some questions for further thought / discussion

- Do you have the intuition that an understanding / explanation of consciousness will involve understanding / explaining qualia or P consciousness?
- If so, then what has cognitive science shown us about consciousness thus far?
- If not, then should ‘Phenomenal consciousness’ go the way of vial spirits (i.e., be eliminated from the subject matter of science)?

9. Possibility, conceivability, supervenience, and zombies

Possibility, conceivability, supervenience, and zombies

- Possibility (logical, physical, biological)
- Conceivability (seems possible to me for all I know - epistemic possibility)
- Supervenience (baldness supervenes on hair distribution)
- Zombies - (intended to show that consciousness does not supervene on material states)

Possibility

- Think of ‘possibility’ as a space (the space of possibility)
- Different possible worlds (located within that space) are different ways the world might be
- E.g., there is a possible world in which Bush won the last US election
- The actual world that we inhabit - @
- Sometimes people talk about different kinds of possibility
  - Logical
  - Physical (or metaphysical)
  - Biological

- Each of these notions places different constraints on the limitations of the space / the worlds within that space

Logical possibility

- Logical possibility sets the outer limits on what is possible
- Contradictions are necessarily false or logically impossible - which is just to say they are false in all possible worlds
- Tautologies are necessarily true or logically necessary - which is just to say they are true in all possible worlds
  - Logical or mathematical truths are thought to be tautologies (e.g., 1+1=2. Either p or not p.

Physical possibility

- Physical possibility is a subset of logical possibility because it has an additional constraint:
- Not only are contradictions ruled out - but worlds that are inconsistent with the laws of physics at @ (at the actual world) are ruled out
- So while it is logically possible that there are worlds with laws of physics that are different from ours (there is no contradiction in that)...
- It is not physically possible that there are worlds with laws of physics that are different from ours

Possibility and conceivability

- Conjectures in math are either true or false
- If they are true they are necessarily true (there are no worlds in which they are false)
- If they are false they are necessarily false (false in all worlds)
- However, from my point of view it seems to me (in some sense) that it is 'possible' that it be true and also 'possible' that it is false - as I could conceive (in some sense) of the math turning out either way

Conceivability / epistemic possibility

- What ‘seems possible to me given the state of my knowledge’ is what is conceivable to me, or what is epistemically possible for me - given what I know
- Conceivability has to do with what finite minds like ours can imagine (or what we think we can imagine)
- Conceivability seems to be subjective in the sense that different people could conceive or fail to conceive of different things

Possibility

- Possibility has to do with what does or does not follow given certain constraints (e.g., non-contradiction or the laws of physics)
- Possibility is objective. What is and is not possible is independent of what any of us think is possible (e.g., ‘Goldbach’s conjecture is true’ is either necessarily true else it it impossible). So, in some sense, it is not possible that it turn out either way.
- By analogy, let us suppose that on day one God does two (and only two) things:
  - God fixes all the laws of fundamental physics
  - God fixes the nature and distribution of the fundamental units of the fundamental physics
- Question: On day two does God have more work to do? Does he have to fix the nature and distribution of the fundamental units of chemistry?
- Or does the chemistry just ‘fall out of’ (so to speak) the physics?

Supervenience

- To say that A supervenes on B is to say: There cannot be a change in A without a change in B
  - E.g., to say that ‘baldness supervenes on hair distribution’ is to say that if two people are alike in hair distribution then they are alike in baldness

- It is important that supervenience is an a-symetric relation. To say that A supervenes on B does not rule out a change in B without a change in A
  - E.g., to say that 'baldness supervenes on hair distribution' does not rule out two people being alike in their baldness but different in the way that hair is distributed (people can go bald in different places on their scalp)
- To say 'chemistry supervenes on physics' is to say that two worlds cannot b e alike in physics without being alike in chemistry
  - So that God could have rested on day 2 if he had have chosen to have approached day 1 by fixing the laws of physics and the distribution of physics fundamental particles
  - It is not to rule out the possibility of two worlds that are alike in chemistry but different in physics
- This is similar to multiple realizability

Zombies

- Imagine (conceive of) a world that is a complete physical duplicate of this world (the actual world)
- On this world you have a counter-part (defined as an atom for atom duplicate of you)
- Your counter-part says the things you say and does the things you do...
  - Seems to be an A-consciousness duplicate of you
- Zombies are defined as physical duplicates of actual people that lack p-consciousness
- The possibility of zombies would mean that p-consciousness does not supervene on physics
- The possibility of zombies would show dualism to be true of p-consciousness
- Materialists counter that while we may think that zombies are possible we are mistaken and they aren't possible at all.
- Materialists say that they are conceivable but not possible and one can't infer possibility from conceivability (e.g., in the case of Goldbach's

conjecture)

- Else they say they don't believe in p-consciousness (eliminativism)
- Next week: More on consciousness

10. The possible world framework, correlation, and identity

Plan

- Last time we briefly looked at some of the ways that consciousness has been operationalized so that we could get underway with a science of consciousness
- Then we turned to the possible worlds framework to try and understand the difference between possibility and conceivability
- This time we will start by using the possible worlds framework to try and understand correlation and numeric identity (p=p)
- Then we will return to operationalizations, and the issue of whether discovering neural correlates of consciousness shows us that consciousness is one and the same as the physical correlates

The possible worlds framework

- Braddon-Mitchell and Jackson, and Chalmers say to think of possible worlds as 'universes'
- Modal realists think that possible worlds are objectively existing and concrete (though spatio-temporally and causally isolated from this world)
- Other theorists think that possible worlds are best understood as something along the lines of sets of 'maximally complete sentences / propositions'

Possible worlds

- Fictional worlds are incomplete insofar as there are truth value gaps
  - E.g., ‘Cinderella had 10,000 hairs on her head when she put on the shoe that fit’
- Possible worlds are maximally complete insofar as there are no truth value gaps
  - Either because there is a fact about the world (on a modal realist view)
  - Or because worlds are constructed by stipulating (consistent) truth values for sentences / propositions (on the view that they are sets of sentences / propositions)

Correlation

- x and y are correlated in the actual world if whenever x occurs y occurs (and vice versa)
- Correlations can be contingent, however
- This is just to say that while x and y might be correlated in the actual world it might be possible that they not be correlated
- This is just to say that there are possible worlds (or there is a non-contradictory set of sentences describing a situation or world) in which they aren’t correlated

Identity

- x and y are numerically identical if there is one object (substance, property etc) rather than two
- An object (substance, property etc) is numerically identical to itself
  - p=p, or brain state x = brain state x
- If there is no correlation between x and y then x and y cannot be numerically identical (one and the same object, property, etc)
- Leibniz law describes this:

  - If x has a property that y lacks (or vice versa) then x does not = y

Contingency of correlation, necessity of identity

- Correlations may be contingent
  - If x and y are actually correlated it may be possible that they not be
- Identities are necessary, however
  - If x and y are numerically identical then they are in all possible worlds
- This is because an object is always identical to itself (p necessarily = p, brain state b necessarily = brain state b)
- So, while correlation in the actual world is necessary for identity it is not sufficient
- That is to say that if there is not a correlation there cannot be an identity but if there is a correlation this is not sufficient to establish identity

Informational value

- To say that ‘p=p’ or ‘brain state b = brain state b’ seems uninformative
- To say that ‘brain state b = mental state m’ seems informative, however
  - Informativeness seems to do with the state of our knowledge
  - Conceivability was relative to the state of our knowledge
- But it is that we can conceive of things turning out either way rather than it being possible for things to turn out either way
- If the identity holds in the actual world it holds in all possible worlds (it is necessary)

Gold

- Gold = 79 protons in the nucleus of the atoms
- if the above identity claim is true then it holds in the actual world and in all possible worlds
  - If we were able to remove a proton from the nucleus of the atoms of a sample then we would have transmuted the substance from gold to something else
  - Similarly, if there was a possible world in which the yellowy malleable valuable stuff turned out to have a different number of protons in the nucleus then that substance would not be gold

Water

- Water = H_2O
- If the above identity claim is true then it holds in the actual world and in all possible worlds
  - So, if the colorless, odorless stuff that falls from the skies and fills the lakes, the drinkable potable stuff is XYZ then it is not water

Neural correlates of consciousness

- Thus both materialists and dualists can be interested in discovering the neural correlates of consciousness
  - Identity theorists think that the discovery of neural correlates is a discovery of the identity of conscious states
  - Dualists think that the discovery of neural correlates is nothing more than that
  - Functionalists think that the discovery of neural correlates that fill the functional role isn't a discovery of the identity of conscious states (because they identify conscious states with the role being filled rather than with the filler of the role)

11. Operationalization, neural correlates of consciousness, qualitative experience

Anxiety

- Suppose you want to study something along the lines of:
  - how x affects anxiety
- You will need to start out by operationalizing (providing a measure of) anxiety
  - Nailbiting
  - Fidgeting
  - Physiological arousal
  - Verbal report (e.g., ‘I feel anxious’)
- Since you want your experiment to be replicable you want inter-rater reliability (other scientists to agree with your rating or scoring of the presence or absence of anxiety or symptoms of anxiety)

Consciousness

- Suppose you want to study something like ‘the effects of x on conscious experience’
- You will similarly need to start out by operationalizing (providing a measure of) consciousness
  - Verbal report

  - Effects on behaviour
  - Physiological arousal
  - Neural activity

Problems with operational definitions

- It can be unclear how well operationalizations measure what they are intended to measure
  - E.g., People bite their nails for reasons other than anxiety and people with anxiety may not bite their nails
- If different research studies or different research groups operationalize differently then it can be unclear how much they are really measuring or studying or talking about the same thing
  - E.g., physiological measures of anxiety in rats, reports of feeling anxious in people etc

Indirect measures

- Behaviourists objected to studying conscious states because conscious states (in others) are unobservable
- They thought that in order to to science we needed to refocus on behaviour (including verbal report)
  - Other sciences (e.g., Physics) use indirect measures, however
- We do this in practice, however, by focusing on measures of behaviour and verbal report
  - It is just that we take the behavioural measures to provide indirect evidence of conscious states

Neural correlates

- The search is on to find the neural correlates of consciousness (NCC's)
- We have already considered that it is problematic to conclude an identity from the discovery of correlation, however

- Even if the correlates of consciousness are found there is more work to do to show the relation to be one of identity

Zombies

- Last time we considered the possibility of zombies (as opposed to conceivability of them) would undermine materialism
- Some people strongly have the intuition that there is no logical contradiction in there being a physical duplicate of this world that is not a phenomenological duplicate
    - This is just to say that what it is like to be your counter-part is what it is like for you when you are in a dreamless sleep

Spectrum inverts

- The possibility (as opposed to merely the conceivability) of spectrum inverts would similarly undermine materialism
- When you look at an object that you have learned to call 'red' you have a qualitative experience with a certain character (p)red (for phenomenal red)
- When you look at an object that you have learned to call 'green' you have a qualitative experience with a particular or peculiar phenomenal character (p)green
- A spectrum invert has inverted qualitative spectrum experiences to you
- When your spectrum inverted twin looks at things they have learned to call 'red' they have the (p)green experience that you have in response to viewing things that are green

Impasse

- If it is possible that a physical duplicate world could have a counter-part of you with either:
    - No conscious experience
    - Inverted conscious experience

- Then this shows that whether a being has conscious experience or not (and the character of the conscious experience that it has) are not determined by, do not supervene on, and cannot be identified with material states
- Materialists say that inverted spectra and the lack of phenomenal consciousness in a physical duplicate world is not possible
- This is because they think that phenomenal experience and the character of the phenomenal experience are logically determined by the state of the physical world and the physical laws
- Dualists deny this. They say that there is no logical contradiction in a physical duplicate world that contains either zombies (with no p consciousness at all) or spectrum inverts

Proceeding with the science

- Even if you think that zombies or spectrum inverts are possible (that there is no contradiction)
- You still might think that there is a point to learning about the actual neural correlates of consciousness (or learning more about the functional roles that conscious states play in the actual world)
- it is just that you think that we aren't entitled to infer the identity from a correlation

12. Symbol systems, intelligence, the Turing test

Symbol systems

- Newell and Simon (pioneers of artificial intelligence) say that a Symbol system is:
  - A physical device that contains a set of interpretable and combinable items (symbols)
  - And a set of processes that can operate on them (copying, conjoining etc)
  - And that these are necessary and sufficient conditions for intelligent action
- The meaning (content) of the symbol is meant to be determined by its place in the network (of inputs, other states, and outputs)
- Meant to be a case of the semantics (meaning / content) being determined by the syntax (place in the network or processes operating on the symbols)
- Like how the states of the soda machine table (0, 1, 2, 3) were defined in relation to inputs, other internal states, and outputs
- The plan is to use a symbolic code to store long term knowledge (a knowledge database)
- Intelligence is then the ability to successfully search the database to find a solution to a given problem

AI symbol system programmes

- Restaurant script
- SOAR

Transparency of symbols

- We think about things like trees, cats, colours etc
- A ‘transparent symbol’ is a symbol with content that is familiar to us
- The machine table helps make sense of symbols that have content that are not transparent to us
  - Cognitive psychology?
  - Neuroscience?

Intelligence

- We take some behaviours to indicate intelligence
  - Playing chess
  - Solving equations or ‘real world’ problems
  - Conversing in language
- Intelligence seems (intuitively) to have something to do with sophisticated cognitive processing

Can a machine think?

- There has been much controversy over whether machines could (one day) think
- There was much controversy over how we would know whether a machine really was thinking or not
- Alan Turing proposed what has come to be known as the Turing test of artificial intelligence
- Three independent judges converse via tele-terminal. They can ask whatever questions they like

  - They might be conversing with a human
  - They might be conversing with a computer
- If the computer fools two out of three judges then the computer wins
  - The computer deserves to be regarded as a genuine thinker
  - So then the issue becomes ‘can a machine pass the Turing test?’

Eliza

- Eliza was developed as a model of a Rogerian Psychotherapist
- The programme takes sentences of English as inputs and transforms them into outputs as Rogerian Psychotherapy style sentences of English
- Transcripts of conversations show it to do fairly well at times
- It has a number of fall-back phrases
  - Hmm. Interesting.
  - Tell me more.
- Check out a version of the programme online!

Parry

- Parry was developed as a model of Paranoid Schizophrenia
- The programme takes sentences of English as inputs and transforms them into English outputs (Paranoid Schizophrenia style)
- Transcripts show it does pretty well sometimes. Also makes frequent uses of fall-back phrases
  - The Mafia are after me!
  - Are you thinking of hurting me?

Turing test

- 2 out of 3 psychiatrists thought that they were conversing with a person with paranoid schizophrenia after a conversation with Parry

    - Parry passed!

- A problem with these programmes is that they aren't 'normal' people so they make use of fall-back phrases

- So maybe Parry did not really pass the test

- The judges may have also underestimated the present state of AI and not asked probing enough questions

- The psychiatrists were also ethically limited because of the possibility that their questioning might genuinely further disturb an actual psychiatric patient

- 'odd' answers were attributed to the Rogerian orientation or to the paranoia

- There was a joke that it told us more about the intelligence of psychiatrists than about the intelligence of artificial intelligence programmes

Can machines exhibit intelligence?

- The invention of the pocket calculator seemed to debunk solving equations as being an indicator of genuine intelligence

    - Pocket calculators can calculate but they don't seem to be very intelligent or to be *thinking* or *understanding* what they are doing

        - A good model of a person obsessed with numbers?

- It might be that it is not enough to *behave* intelligently (text output as behaviour)

    - Once we find out how something does it the explanation can be 'debunking'

- A criticism of AI is that while computers might mimic behaviour the way in which they do it seems to be very different (their architecture is very different from ours)

- Another criticism of AI is that no amount of state transitions can give us genuine intelligence

13. Objections to symbol systems

Plan

- Searle's 'Chinese Room' objection
- Block's 'Population of China' objection
- Argument from the 'Fluidity of Everyday Coping'
- Microfunctionalism

The Chinese Room

- It is a thought experiment (use your imagination):
  - There is a person in a room who speaks only English
  - Papers come into the room through a mail slot and they have marks on them
  - The person has a book that tells them if there are certain marks they are to write down certain other marks
  - the person then posts the results back through the mail slot
- The papers that come in are actually questions that have been written down in Chinese
- The book is a translation manual that provides answers in Chinese to Questions in Chinese
- The papers that go out have answers, then, to the questions in Chinese
- Does the person in the room understand Chinese?

- That was supposed to be rhetorical with a resounding NO.
- Similarly (the argument goes) a computer that manipulates symbols according to rules doesn't *understand* the content of the symbols / meanings
- One response:
  - The relevant analogy isn't between the person in the room and a symbol manipulating computer
  - The relevant analogy is between the person + the translation manual and a symbol manipulating computer
  - While the person in the room might not understand Chinese, the person + the translation manual do

The population of China

- Another thought experiment
  - Imagine that we take the population of China
  - We get them to implement the functional profile of a mental state (e.g., the belief that the sun is hot)
  - That means to say:
    - Give them letters or other formal symbols
    - Instruct them to pass the letters to each other according to certain rules
  - Does the population of China exhibit the relevant mental state?
  - Is that how to induce a belief?
  - If you have the intuition that they don't, then you might well think that no amount of mere symbol manipulation is sufficient for mentality
  - If you have the intuition that symbol manipulation is sufficient for mentality then you might be inclined to think that the population of China could be manipulated in this manner or way

Fluidity of everyday coping

- What would an AI do if it discovered a Martian in the kitchen?
  - Add more to the knowledge base
  - Add a more powerful inference engine to the knowledge base
  - The point is that it would do more of the same

Micro-functionalism

- Clarke says that we might be able to fix the 'finer details' of the internal state transitions such that they can't b e replicated by the population of China
  - Might lose multiple realizability if the details get too fine?
  - Ad hoc?

14. Folk psychology, the life world, stances

Folk Psychology

- Folk psychology involves ascribing mental states such as belief, desire, hope, fear, etc
- Mental states are thought to be propositional in structure (subject S believes that p (believes that some proposition p is true), desires that q)
- We ascribe these states (at least in part) for the purposes of prediction and explanation of behaviour

Fodor

- Focuses on the success of folk psychology because:
- Representational theory of mind (RTM)
    1. Propositional attitudes pick out computational relations to internal representations
    2. Mental processes are causal processes that involve transitions between internal representations

Churchland

- Focuses on the limitations of folk psychology
- It works only sometimes

  - Not good for explaning sleep, mental illness, neurological illness etc
- Folk theories typically have a bad fate as scientific theories
  - Folk biology and folk physics did not fare well
  - Folk psychology has not developed or progressed
- Folk psychology does not fit with science (neuroscience or physics)

Folk psychology

- They both agree that folk psychology is committed to mental states being internal states that cause behaviour
  - Fodor - there are such states
  - Churchland - there are no such states

Abstracta

- Physicists posit centres of gravity
- They then use this object (a centre of gravity) to predict the behaviour of physical objects
- Are there really such things as centres of gravity?
- Before we can say whether there really are such things we need to know a bit more about what they are supposed to be
- If a centre of gravity is a point mass then there isn't any such thing (it would be a useful fiction)
- If a centre of gravity is a vector sum that acts through a point then there is (they are perfectly real)
- Center of the population of the United States (there may be no particular person at that point)
- Dennett's lost sock centre
- Abstract but not useful - real?
- Reality comes cheap for Dennett - but he professes to be less interested in 'reality' and more interested in 'utility'

  - Headline - Scientists discover that left handed people don't really have beliefs!
  - Headline - Scientists discover that people with diabetes don't really have desires!
- If there weren't inner states that played the right causal role this wouldn't undermine folk psychology
- We need to focus on 'the light' (what is visible to us) - behaviour
- What vindicates folk psychology is the predictive leverage that we get from it

Conway's Life simulation

- https://conwaylife.com/
  - A grid of cells where each cell can be in one of only two conditions on or off
  - Time is discrete (advances in ticks)
  - One law of physics: If 2 neighbours are on stay same, if three neighbours are on then on, else off
- The physical stance
  - The ontology of the 'physical stance' consists in cells that are either on or off
  - the state of the life world at the next tick can be predicted completely by the state of the life world at the previous tick together with the laws of physics
  - 100 per cent accuracy
- The design stance
  - If you run the programme so that (for all you know) time in continuous
  - A new ontology emerges
  - You can see for yourself that there are objects that persist through time

    - A block of 4 cells that are on will stay on, or persist though time
    - A flasher will flash indefinately
    - A glider is an object that moves through space
- We can make true generalisations about the behaviour of these objects. For example:
    - 'An eater can eat a glider in four generations [ticks]. Whatever is being consumed, the basic process is the same. A bridge forms between the eater and it's prey. In the next region the bridge region dies from over-population, taking a bite out of both eater and prey. The eater then repairs itself. The prey usually cannot. If the remainder of the prey dies out as with the glider, the prey is consumed'.
- The predictive leverage that we get from the design stance is less than 100 per cent
    - Provided that nothing else encroaches
- Do the objects from the design stance 'really exist'?
- Temptation is to say: Sure they do – look for yourself!
- Can the objects from the design stance be identified with physical cells?
- They are multiply realised by them (consider gliders)
- While we can predict the next state of the lifeworld with complete accuracy from the physical stance it doesn't seem able to capture some things:
    - 'An eater can eat a glider in four generations'
- The temptation is to say that if one didn't see the ontology of the design stance (the flashers etc) then one would be missing something that was really there
- In the actual world we see the tables and chairs
- Beliefs and desires?

15. Intentional stance, indeterminacy, real patterns, true believers

The intentional stance

- The design stance involves our seeing objects that persist that seem to move through time and space
    - Insofar as we didn't view the lifeworld from the design stance we were missing something that was really there
- The design stance works well for artefacts
    - Clocks and thermometers behave how well they are designed to behave
- The intentional stance involves viewing the behaviour of objects as the rational product of mental states
    - Look to where the 'light' is (behaviour)
- An intentional object will act so as to satisfy their strongest desire on the assumption that their beliefs are true
- An intentional object is motivated to revise their beliefs towards truth
- An intentional object is happy when they get what they desire
- We use platitudes like these to predict and explain behaviour from the intentional stance
- It works other things being equal
    - Fails for neurological breakdowns, quirky environments etc

  - Similar to the ‘provided that nothing encroaches’ clause

- Dennett thinks these generalizations work fairly well for people because evolution eliminated those who weren’t rational

Real patterns

- In his early works many thought that he was a ‘fictionalist’ about mental states
  - ‘They are, strictly speaking, fictions’
- Or an ‘instrumentalists’ about mental states
  - They are ‘useful tools’ but there really isn’t any such thing
- Or an ‘irrealist’ about mental states
  - They don’t really exist
- In ‘Real patterns’ (one of his later works) he tries to focus on the realist aspects of his view
  - Utility of abstract objects (including patterns)
  - Mind-independence of patterns (if we don’t see them we are missing something that is really there)
  - Many think that his ‘mild realism’ is unstable and that it lapses back into realism
- Example of bar code
  - Same pattern with different noise ratios
- Patterns are objective
  - ‘Pattern’ is defined in terms of compression (e.g., from a pixel by pixel specification)
  - Someone can fail to see a pattern that is really there
  - We probably fail to see patterns that are apparent to other organisms

Indeterminacy

- Dennett does think that there could be a case where different people see different patterns
  - 60% pattern, 40% noise
  - 70% pattern, 30% noise
  - If both play the odds properly both will get rich
- In this case there is no further fact of the matter as to which pattern is real or even 'most real'
- Similarly, there may be no further fact of the matter whether a person 'really believes' p or whether they 'really believe' q instead
- There can be indeterminacy in concrete objects too
  - How many grains of sand make a heap?
  - When is a person bald?
  - When does a hill become a mountain?
- So if it can be indeterminate whether someone really believes p or whether they really believe q instead that doesn't necessarily undermine the reality of belief

True believers

- What makes it true that an object has the belief that p?
- When we adopt the intentional stance towards an object (view it as a rational agent with beliefs and desires) then there is predictive leverage to be had by ascribing the belief that p to it
- There is a concern that the theory is too promiscuous in what gets to count as having mental states
  - What about a chair staying still because it feels like it?
  - What about a thermostat having beliefs and preferences about room temperature?
  - What about a chess playing computer that desires to get it's queen out early?

Indeterminacy of true believers

- There may be indeterminacy whether an object really is an intentional object
  - Continuum (in some sense) between us and animals
  - Continuum in our evolutionary history
  - Continuum in our individual development
  - Continuum in artificial intelligence programmes
- But that doesn't necessarily undermine the reality of beliefs and desires

Predictive leverage

- Perhaps we are only justified in adopting the intentional stance when there is predictive leverage that can be had by no other method
- Still problems with pattern / noise trade-off
- Maybe it is okay that there is a degree of indeterminacy in who the true believers are or in what it is that an intentional object believes

Summary

- In 'Real patterns' Dennett attempts to focus on the reality (objectivity, mind-independence) of patterns specified in terms of compressibility
- Two different patterns can be equally real insofar as the people who see different patterns can both get rich playing the odds
- Real patterns in behaviour emerge when we adopt the 'intentional stance' towards an object - just like how real objects emerge in the life-world when we adopt the 'design stance'
- True believers are objects whose behaviour is successfully predicted from the intentional stance
- The predictive leverage is to be had by no other method

16. Connectionism and symbol systems

Dennett (again)

- Intentional action
  - A way of viewing behaviour (from the intentional stance) such that it is the product of a rational believer
- Examples?
  - Hailing a taxi
  - Requesting to speak
  - Ordering 500 shares in General Motors
- I can order 500 shares in General Motors by, for example:
  - Picking up the phone with my left hand
  - Picking up the phone with my right hand
  - Emailing my bank
  - Emailing General Motors
  - Going to see my stockbroker...
- Someone who fails to see that these are all ways of ordering shares is missing a real pattern that is out there in the world
- If I know that you desire to make money and I know that you believe the value of GM shares are about to go up then I can predict (from the intentional stance) that you will order shares in GM

- But I can't predict whether you will pick up the phone with your left hand or whether you will email...

Where we are at

- Introduced folk psychology and the idea that mental states are propositional attitudes
- Folk psychology, autism, and modularity
- The nature of mental sates
  - Can computers have whatever it is that we have that allows us to have mental states?
- What would that computer be like?
  - What is it that we have that allows us to have mental states?

Symbol systems

- Based on the notion that the structure of language, logic, and thought is the same
- That is to say they are all propositional in nature
  - Meaningful units (symbols or symbols structured into bigger meaningful units - propositions)
  - Mental states (beliefs, desires, emotions etc) are attitudes (relations) to these propositional contents
- Symbols (meaningful units)
  - E.g., John. Cat.
- Rules of combination so they can create bigger meaningful units
  - John likes the cat - meaningful
  - The cat likes John - meaningful
  - Cat Jon the likes - not meaningful
- Rules of inference / deduction / state transition

  - If George is a cat and all cats are mammals then George is a mammal
  - If input 50c then goto state 2
- Symbol string encoding
  - Body of declarative statements written in formal notation based on the structure of language and logic (e.g., LISP programming language)
- Serial, feed-forward processing
  - One processing stream
  - Feeds sequentially forwards from input to internal state to internal state to output
- In a discrete state (symbol or proposition) then transition to another discrete state

Connectionist systems

- Also known as:
  - Parallel distributed processing (PDP)
  - Artificial neural networks
  - Units are input, hidden, and output
  - Weighted connections
- The number of units, the number of connections, and which units are connected are decided by the architect
- The initial weightings are often randomly set
- The designer then gives the network a series of training cases and the network learns from the training cases
- To begin with the outputs tend not to be the desired ones
- Backwards propagation learning algorithms can then be used to adjust the weight so that it outputs what we want
- With a large number of training cases it eventually gets the outputs we want

- Processing is distributed as activation over different units
  - Symbolic architectures were in discrete states
- Processing occurs in parallel with many connections participating in producing the output
  - Symbolic architectures had only one serial processing stream of state transitions
- Connectionist networks aren't explicitly programmed with a knowledge database and a series of rules for state transitions
  - Symbol systems had knowledge databases and rules programmed into them in symbolic languages

Symbolic example - DECtalk

- A model of grapheme (letter) to phoneme (sound) transition (text to speech)
  - Programme a knowledge database of rules and list the exceptions to the rules

Connectionist example - NETtalk

- Text to speech
- Fix the architecture (the number of units and the weighted connections between the units)
- Then use a learning algorithm with backwards propagation
- Feed it a large set of training cases
- The outputs were initially not what we desired
- Semi-recognisable words and syllable structure emerged
- Then it got pretty good
- NETtalk had
  - Trained input of 7 letters
  - 7 groups of input units

- Each group comprising 29 individual units whose overall activation specified one letter
- 80 hidden units, 26 output units, and 18,829 weighted connections

17. Connectionism: Features and problems

Summary

- Features
  - Neurology
  - Behaviour
  - Learning
- Problems
  - Folk psychology
  - Systematicity
  - Commonality
  - Post-training analysis
  - Biology

Features - neurology

- Connectionst networks seem more neurologically plausible than symbol systems
  - Units bear a striking resemblance to neurones
  - Connections bear a striking resemblance to axon - dendrite connections between neurones

  - Distributed activation (parallel processing) seems more neurobiologically plausible than serial feed-forward

Features - behaviour

- Neural networks seem to be good at the things we are good at
  - Generalising to new cases
  - Recognising objects and patterns
- Neural nets seem to be bad at things we are bad at
  - Sequential logic or mathematical derivation

Features - learning

- Networks seem to recapitulate our learning
  - Similar sorts of errors to human infants
  - Eventually gets generalisations right
- Networks seem to recapitulate the way we unlearn
  - More damage results in worse performance (graceful degradation)
  - Eventually gets generalisations wrong

Problems - folk psychology

- Our common sense folk psychological intuition:
  - Mental states are functionally discrete, semantically interpretable, inner states that are symbolic in nature
  - They play a causal role in inference and in behaviour
- The neurologically plausible feature of PDP networks (distributed rather than discrete and parallel rather than sequential) seems to sit badly with our folk-psychological intuition about mental states

Problems - systematicity

- Fodor - thought is systematic so internal representations are structured
- Connectionist models lack structured neural representations
- So connectionist models aren't good models for thought
- To know a language involves knowing the parts and how they fit together
- E.g., if you can think 'the Cat loves John' you can think 'John loves the cat'
  - Maybe connectionist architectures can support this
  - Or maybe thought derives it's systematicity from language - and not the other way around

Problems - commonality

- Could have a network that can store 16 propositions and the same network updated to store 17
- There might be no overlap in node activation or weightings
- Doesn't seem to be anything in common between the networks (but they have common propositional contents)

Problems - post-training analysis

- How do we figure out what kinds of representations the network has acquired?
- Cluster analysis (statistical technique)
- Damage
- Microstructural content e.g., black cat in the visual field with minor variations for different orientation
- Panther and cat aren't semantically overlapping (though black cat and black panther are) but might have overlapping features
- Sale and sail should overlap (same output) whereas pint and hint (despite substantial letter overlap) are quite different phonetically

- Cluster analysis can help us recover more traditional symbolic contents from connectionist architectures
- Concern that connectionist architectures might just be fancy symbol systems
- Clarke (textbook) thinks that adding temporality prevents this
- Three responses:
    1. Look harder to cluster profiles for discrete symbolic content (Clarke)
    2. Folk psychology does not commit us to discrete symbolic content (Dennett)
    3. So much the worse for folk-psychology for committing us to discrete symbolic content (Churchland's eliminativism)

Problems - biology

- Very simple models
    - Limited perceptually (typically to only one domain e.g., verbs, letters, pictures)
    - Limited behaviourally (typically not actual motor action)
- More Neurology data might be important as we might be able to join them up