Philosophy and Cognitive Science

Topic 1: Introduction to Cognitive Science

What is Cognitive Science?

- The scientific study of cognition, or mental states
- Often uses computational models
- Seeks naturalistic theories to explain mental states, their nature, meaning and function, and their connection to behaviour
- Cognitive Science is Interdisciplinary. A research program involving researchers from many
- different disciplines including: psychology, Artificial Intelligence (AI), Linguistics, Evolutionary Psychology, Neuroscience, Philosophy

Types of Mental States

- Perceptions: seeing, hearing, smelling
- Sensations: pains, itches, hungers
- Emotions: love, gratitude, happiness
- Thoughts: beliefs, desires, memories

Features of mental states

- Causally efficacious: They do things to the world (e.g. lead to creation of a building), to our bodies (e.g. make me type what I am typing), and to other mental states (e.g. my feeling of thirst causes me to wonder where I left my water bottle)
- Conscious: phenomenal quality, there is something it is like to be in them
- Contentful: 'about' something, often something relating to the external world
- Dependent for existence on the brain

Two Types of Computation

- Cognitive science seeks naturalistic theories to explain mental states, their nature, meaning and function, and their connection to behaviour. These explanations are often computational, in terms of representations and the causal interactions between them.
- Two research programs have dominated. They hold different views on the nature of the computation involved.
- Classicism or good old fashioned AI: holds that the mind is a digital computer that processes symbols according to syntactic based rules. In these sorts of systems it is the rules (the program) that determines how the computation proceeds
- Connectionism: holds that the mind is an analog computer that uses representations that resemble the world in some direct way. In these sorts of systems it is the nature of the representing vehicles that determines how the computation proceeds (the computational model behaves like the world because it is structured in the same way as the world)

Innateness

- To what extent are mental states innate or acquired, and what is the interplay between these processes?
- Related question: the relative contribution of nature or nurture to behaviour
- Twin studies provide an excellent design to shed light here

Folk Psychology

- Folk psychology is our everyday "folk" understanding of psychology. This is distinct from scientific psychology
- Cognitive scientists are interested in examining our folk theories of minds
- Folk psychology understands and explains behaviour in terms of mental states
- Example of folk psychological reasoning: John went into the kitchen because he was hungry
- Crucial to such reasoning is a 'theory of mind', or the ability to attribute mental states like beliefs and desires to people and understand their behaviour accordingly
- People with autism seem to have a much less developed theory of mind

Meaning

- How do my thoughts, perceptions, sensations, emotions get to be about things in the world and my body?
- Twin Earth thought experiment of meaning – do Earthians and Twin-Earthians mean the same thing when they say 'water'? Since they are referring to different substances it seems not, in which case it seems that 'meaning aint in the head'

Topic 2: Philosophy of Mind

Cartesian Dualism

- Descartes concluded that he was essentially a thinking thing, housed in a physical body, like a 'pilot' in a ship (or as was later stated by Gilbert Ryle, a 'ghost in the machine')
- Thus, he believed that he was a mental substance, a thing that did not depend for its existence on the physical world
- Four features of Cartesian egos that distinguish them from physical things
    - Subjectively transparent: All and only mental things are accessible to consciousness
    - Unity: The soul is not divisible. Each conscious mind has only one point of view on the world. All conscious experiences are unified in one conscious subject
    - Privacy: I cannot have your conscious experiences and you cannot have mine
    - Authority: I have first-person authority over my mental states. I can't be wrong about the experiences that I am having and only I have this sort of access

Problems with Dualism

- Unconscious mental states: e.g. Freudian slip. I have unconscious, repressed mental states that can slip into conscious thoughts and affect behaviour
- Disunified consciousness: case of split brain patients
- Privacy?: Cases of conjoined twins with shared neural pathway, have experiences that they attribute to the other twin. Can access an experience that is not theirs

The Problem of Other Minds

- How can anyone be sure that others – other humans, animals, perhaps even machines – are thinking, feeling, conscious beings? Perhaps everyone else is a zombie, or an automaton
- Yet we believe there are other minds like our own by making generalisations from behaviour, our own and theirs: I have a mind and it is responsible for my behaviour, other people behave like me, so they must also have minds that are responsible for their behaviour
- This isn't really induction though, as we don't have direct observational access to both events; it is more an inference to the best explanation

The Argument from Analogy

- The traditional solution to the problem of other minds has been the analogical inference to other minds
- Other human beings are very like me. They behave much as I do in similar circumstances and they are made of the same stuff. When I burn myself it hurts and I cry out and wince. When other people are burned they do the same. I can thus infer that they are in pain too

Problems with the Argument

- One problem with this argument is that it seems impossible to test the conclusion, which makes it a rather problematic inference, although arguably not uniquely so in science
- Another rebuttal is that this 'inference' is a generalization based on one case only and therefore fatally unsound
- It would be like a Martian inferring that all humans are babies based on an encounter with a single human (a baby one)
- It seems that we acquire our concepts of mental states via our interactions with other people (e.g. the case of Genie, who was kept bound and socially isolated all her young life till the age of 13, at this point she was virtually mute and lacked a folk psychology)
- This suggests that the way we grasp mental state terms like 'sensation', 'thought', 'belief', 'itch', 'sound', 'envy'...hundreds of others, depends on exposure to social life
- Those defenders of the analogical inference who accept that it is an inference based on one case, obviously insist also that the analogical/hypothetic inference remains a sound inference. They claim that what is needed is a causal link between events, in this case between mental events and other events. The demand for more than one case is legitimate where more than one case is needed to establish a causal link between events. However, one case will be enough when that one case can establish that the link holds. It is argued that we have sufficient resources available from our own case to establish that the relevant causal link holds (Hyslop and Jackson 1972)

Behaviourism

- Behaviourism was a movement against the Cartesian picture, seeing it as unscientific, and seeing it as unable to adequately deal with the other minds problem
- In addition there was a claim that science could not study what was essentially private, or what could not be systematically and objectively analysed
- Behaviourists viewed the alleged internal aspects of mind as like an impenetrable black box; thus, not the kind of thing one could study

- Methodological behaviourism: a position in scientific psychology which was neutral as to whether mental states exist, but claims that all that can be studied is behaviour
- Philosophical behaviourism: holds that mental states are defined by behaviour and dispositions to behave - there is nothing to the mind beyond behaviour
- Mental state talk does not refer to inner states at all, but to observable bodily movements or the potential for them
- For example, to be in pain is to exhibit pain-behaviour or be disposed to exhibit that behaviour
- Behaviourists thus view mental state terms (in the potential cases) in the same way dispositional properties are viewed
- Examples: The analysis of 'the glass is fragile' is 'if the glass were dropped it would shatter'; 'John is hungry' means 'If John had access to the fridge he'd consume the food in it'

Problems with Behaviourism

- Unsatisfying: Behaviourism is very hard to believe. A famous behaviourist joke bears this out: two behaviourists meet in the lift one day, the first one says 'You're well today, how am I?'
- Behaviour is Underspecified: Ordinarily, we think that people can be in the same mental state, but act very different
- Problem of Deceit: People can lie, deceive and engage in actions that are inconsistent with their beliefs (e.g. pretend to like a person they really don't). Under the behaviourist notion if a person always behaved as if they liked you, then they really do like you, even if they actually hate you
- Qualia: Whether I'm in pain, or having the "reddish" kind of visual experience I normally have when I look at a ripe tomato, seems to a matter of what's going on in me right now, which I can be directly aware of. It doesn't seem to be a matter of how I would act in certain hypothetical situations

Identity Theory

- Around the mid 1950s philosophers and scientists began toying with the idea that the mind is the brain: all mental state types are instances of brain state types (though not the reverse)
- Some evidence:
    1. fMRI scans reveal brain activity that is correlated with patient reports of cognition
    2. brain damage cases are correlated with cognitive failures or personality changes
    3. brain development is correlated with mental development
    4. the cessation of brain activity is correlated with the cessation of consciousness
- This could be compared to past example of inter-theoretic conceptual reductions in the history of science, such as lightning = discharge of electrons, heat = mean kinetic energy, light = electromagnetic waves and sound = compression waves through air
- A somewhat weaker form of identity theory denies that every unique brain state = unique mental state. Instead, they argue that every instance of a particular mental state is also an instance of a particular physical brain state
- This means that you could have similar or even identical mental states with different brain states, so long as they fit into the same classes of states

Problems with the Identity Theory

- Multiple Realizability: mental states can be realized in multiple kinds of systems, not just brains
- Mary Argument: Mary is a colorblind neuroscientist who knows knows all physical facts about colour, but can still 'learn something' by actually seeing colour
- Leibnitz Rule: two things are the same only if they share all the same properties. The mind and the brain do not share all the same properties (e.g. qualia, consciousness), so they cannot be the same thing

Functionalism

- Functionalism is the view that mental states like pain occupy a functional role within the organism in which they occur: it is behaviourism plus mental states
- A pain is an internal mental state that causes me to behave in pain ways and to have other pain related mental states
- For example, 'pain' would be defined as the mental state that followed from physical harm to the body, and causes mental states concerning the desire to relieve the pain, and behaviours such as wincing, crying, etc
- Functionalists do not deny mental states, but argue that they play a causal role of mediating between bodily inputs and bodily outputs
- Functionalism is a causal theory of the mind; it says that mental states occupy functional (or causal) roles, and so in this sense fits well with folk psychology
- The one problem it does not address, however, is consciousness – the problem of what it is like to experience the world from a point of view

Computational Theory of Mind

- The attempt to understand how minds work according to computational procedures
- The basic idea is that the human brain performs transformations of symbolic representations according to syntactically-specified rules
- Syntactic properties of a representing vehicle are those it has in virtue of its material structure - the physical properties of words and sentences, e.g., that it has a certain shape
- Semantic properties of a representing vehicle are those it has in virtue of the information it conveys - the meaning properties
- Computers are semantic engines which operate in accordance with the meanings of their representations. They are able to do this by manipulating symbols on the basis of their syntactic properties. They are designed specifically so that their operations maintain this relation between syntax and semantics.
- The beauty of computation is that 'if you take care of the syntax, the semantics will take care of itself' (Haugeland, 1985, p.106). Each symbol has a unique content, and the rules are designed so that they respect the meaning of the symbols. So the system manipulates the symbols in a way that respects their meaning.
- But the computational processing is driven not by the semantic properties of the system but by its syntactic ones. We just need to build a machine that manipulates symbols according to their shape. By programming it with the right rules we get it to behave as if it were processing them according to their meaning. Digital computers are syntactic engines that behave like semantic engines.

Strong AI

- From Turing: “any cognitive process may be computed simply through the manipulation of symbols”
- According to strong AI the appropriately programmed computer really is a mind. It can “be literally said to understand and have other cognitive states”.
- A criticism of AI is that while computers might mimic behavior they don’t seem to do it the way we do it (their architecture is very different from ours)
- Another criticism of AI is that no amount of state transitions can give us genuine intelligence

Searle's Chinese Room Argument

- Searle argued that intentionality is a product of causal features of the brain, and that instantiating a computer program is never by itself a sufficient condition of intentionality; no program by itself is sufficient for thinking.
- He justified this by appealing to his Chinese Room argument:
    - There is a person in a room who only speaks English
    - Papers come in through a mail slot and they have marks on them
    - The person has a book that tells them that if there are certain marks they are to write down certain other marks
    - The person then posts the result back through the mail slot
    - The papers that come in have questions that are written in Chinese
    - The book is a translation manual of the that provides answers in Chinese to questions in Chinese
    - The papers that go out have answers to the questions in Chinese
    - Does the person in the room understand Chinese? Searle says no
    - Similarly (the argument goes) a computer that manipulates symbols according to rules doesn’t understand the content of the symbols / meanings
    - Symbol manipulation isn’t sufficient for mentality

Problem of Mental Causation

- The problem of mental causation is a conceptual issue in the philosophy of mind. That problem, in short, is how to account for the common-sense idea that intentional thoughts or intentional mental states are causes of intentional actions
- How do intentional states (mental states which are about something) cause action?
- To the extent that we do not have to go outside human physiology in order to determine the causes of any action, intentional action can be fully causally explained by the existence of these physiological antecedents alone. No mention of mental states need enter into the explanation
- This troubles philosophers because, pre-theoretically, it seems that what is crucial in causing a person to act is their mental states. But, given that physiological facts are sufficient to account for action, mental states appear to be superfluous
- Alternatively, if your mental states were causally isolated from your bodily behaviour, then what goes on in your mind could not explain what you do

Topic 3: Modularity and Theory of Mind

Theory of Mind

- Mindreading is our everyday capacity to attribute mental states to others (and to ourselves)
- This capacity permits us to understand that mental states can be the cause of behaviour and thus plays a central role in our capacity to predict and interpret behaviour (our "folk psychology")
- Folk psychological competence appears essential to successful social interactions and may also play an important role in moral competence
- Normally developing children have acquired key elements of this folk psychology (often called 'theory of mind') by the age of five (e.g. false belief tests, attribution of mental states to others, understanding the other's preferences differ from one's own)

Theories of ToM Acquisition

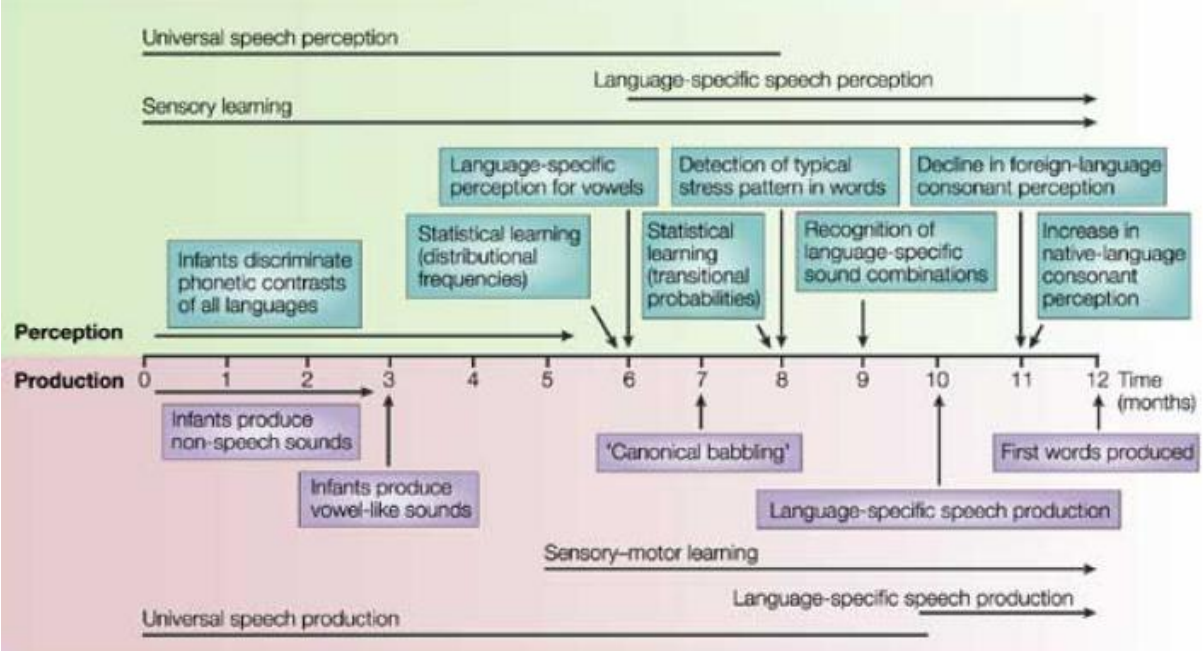
- By Analogy: We know of our own mental states by introspection and we conclude that others have like sensations, thoughts and feelings because they have bodies like ours and "exhibit the acts and outward signs which in my own case I know by experience to be caused by feelings"
- By First Principle: Social intercourse is based on the assumption of mindedness and "it is evident that children are capable of such intercourse long before they can reason"
- Theory Theory: Folk psychological introspection and attribution of mental states to others is an example of theoretical entities posited by us as a theory, which therefore could be demonstrated by science to be incorrect
- Mental Simulation Theory: we represent the mental states and processes of others by mentally simulating them, or generating similar states and processes in ourselves. We can internalize desires and emotions without recognising or conceptualising them

Autism and Theory of Mind

- People with autism have difficulty in mindreading. Probing the autistic child's mindreading deficits may shed light on the cognitive architecture underlying our everyday capacity to predict and explain the behaviour of others
- Specific impairments include lack of empathy, poor communication skills, difficulty in reading body language, lack of creativity and imagination, paucity of pretend play in children, repetitious behaviour
- They do worse on a variety of tests related to mental state attribution including:
    - the mental-physical distinction
    - false belief tasks (notably the Sally-Ann task)
    - "seeing leads to knowing" tests
    - inferring thinking/desire from gaze direction
- Some able autistic people are able to pass false belief tests, although at a much later age than normal children. However it seems unlikely that such individuals acquire knowledge of other minds via the same mechanisms that normals do

Features of Modular Systems

- Modularity of mind is the notion that a mind may, at least in part, be composed of innate neural structures or modules which have distinct evolutionarily-developed functions. Somewhat different definitions of "module" have been proposed by different authors.
- Fodor's characterisation of a module has informed the whole debate. But note that Fodor did not claim that any of these features are necessary, rather sketching characteristic properties.
- Domain specificity: A system is domain specific if it has a restricted subject matter - modules deal with a restricted class of inputs. e.g., colour vision, face and voice recognition. Contrasts with domain general abilities, for example working memory and executive function – which operate over the variety of representations produced by the domain specific systems.
- Mandatory operation: modules operate automatically – that is they are not under conscious control. The system's operations are switched on by presentation of the relevant stimuli. We can't choose not to see in colour or recognise spoken English (or other first language).
- Fast processing
- Shallow outputs: These are non-conceptual outputs. Modules generate information rather than thoughts or beliefs – that would require co-operation from non-modular systems.
- Limited central accessibility: higher levels of processing have limited access to the representations within a module. Internal processing is opaque to introspection – we are only aware of the outputs.
- Informational encapsulation: Module can't be guided by information at higher levels of processing. It cannot access information stored elsewhere. Example is the persistence of visual illusions, e.g., the Muller-Lyer illusion.
- Fixed neural architecture: Modules are realised in dedicated neural architecture
- Characteristic and specific breakdown patterns: Modules can be selectively impaired: that is, damaged or disabled with little or no effect on the operation of other systems. Examples from brain lesion studies include prosopagnosia (impaired face recognition), achromatopsia (colour blindness), alexia and dyslexia all found in otherwise cognitively normal individuals.
- Characteristic ontogenetic pace and sequencing: The system is innate. Modular systems come on line chiefly as a result of a brute causal process of triggering, e.g., language acquisition occurs on the same schedule in all normal individuals



Is there a Theory of Mind Module?

- The claim: "the normal and rapid development of theory-of-mind knowledge depends upon a specialized mechanism that allows the brain to attend to invisible mental states" (Leslie)
- A system is modular if it has enough of the 9 features to an appreciable degree (though Fodor gave special weight to informational encapsulation).
- However, Fodor was providing an analysis of low level systems underlying perception. Can we apply the analysis to higher level systems?
- Massive Modularity is a claim that can be understood in a number of different ways.
    - Weak reading: The mind consists entirely (or almost entirely) of dissociable functional components –rather like an electronic device. So at least some of these components can be damaged without affecting the functionality of the whole.
    - Stronger reading: The mind consists of a great many modular components including higher order cognitive systems.
    - Strongest reading: The modular components would have all of the properties of a Fodor-module.

Evolutionary Psychology Argument for Modularity

- EP is based on the idea that the mind is a set of special purpose modules whose domain specific structure is an adaptation to ancestral environments.
- The focus for these higher order systems is on domain specificity, mandatoriness, localization and/or dissociability and probably speed.
- Cosmides and Tooby's argument:
    - The human mind is a product of natural selection
    - In order to survive and reproduce, our ancestors had to solve a range of problems
    - Since adaptive problems are solved more quickly, efficiently, and reliably by modular (domain specific, mandatory etc) systems than by non modular systems, natural selection would have favoured a massively modular architecture
    - So the human mind is probably massively modular

Further Arguments for ToMM

- Implausibility of theory theory explanation of mindreading: Mental states such as beliefs and desires are highly abstract yet very young infants display implicit knowledge of them whereas they don't display implicit knowledge in similarly abstract domains. '
- Evidence of innateness: characteristic pace and sequencing of the development of theory of mind provides evidence for modularity.
- Evidence of Selective Impairment in Autism: Autistic children are impaired on false belief tasks but perform better than normal 4 year olds on false photograph tasks. This suggests autistic children are not impaired generally in tasks requiring executive function or counterfactual reasoning. The impairment is domain specific, of a cognitive module specialized for representing mental representations. Autistic children are impaired in pretend play and in mind reading.
- Localisation: The temporoparietal junction has been suggested as the location for a domain specific neurocognitive system specialised for ToM, as it is active during false belief but not false photograph tests.

Arguments Against a ToMM

- Pace and sequencing fails cross culturally: Prinz cites evidence of cultural variation. Differences on mindreading tasks are highly correlated with language skills and learning.
- Failure of Localization (fixed neural architecture): Mindreading has been shown to recruit a variety of brain regions including language areas in left frontal cortex, visuospatial areas in right temporoparietal regions, and the amygdala
- Problems with the EP argument: Under this hypothesis the mind would then be rigid, whereas flexibility would have selective advantages in responding to environmental change.
- Mind reading is not like language: No evidence of acquired selective ToM deficit analogous to acquired aphasia. All acquired deficits in ToM also connect with other impairments in lower level or domain general capacities such as joint attention, gaze following etc.
- Defence of theory theory: children *are* actively engaged in theory construction about abstract and invisible causes of physical events, as for example shown in Gopnik's about children's causal explanations and the construction of causal maps.
- Parsimony: Gerrans and Stone say their model is more parsimonious, in which ToM is the emergent outcome of the interaction between numerous domain specific, low level cognitive systems, and other domain general cognitive systems, without requiring its own module.

Topic 4: Delusions

Defining Delusions

- Illusion: is a misperception of something that is actually there
- Hallucination: a perception of something that is not there
- Delusion: a false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary

Major Delusions

- Capgras: Someone close to you has been replaced by an imposter. Accept that they may look the same, but not the same person.
- Cotard: Belief that you are dead. When faced with claim that dead people don't talk and breath thepatient replied, 'well they obviously do'.
- Fregoli: Strangers are people you know in disguise.
- Mirrored-self misidentification: Think reflection is a stranger.
- Thought Insertion: Think that some of your thoughts have been put in your mind by some other agent e.g. the CIA, God.
- Alien Control: Think that someone else is controlling your actions, even in the absence of a medical problem with actual control of your actions.

Categorising Delusions

- Mundane vs Bizarre: Mundane delusions are understandable in terms of folk psychology, such as delusional jealousy or paranoid delusions, and as such can sometimes be true. Bizarre delusions cannot be understood in folk psychological terms.
- Monothematic vs Polythematic: Polythematic delusions extend to more than one theme, where the themes can be interrelated (e.g. persecutory delusions). Monothematic delusions are about one theme and apart from the content of delusion itself, no other (unrelated) bizarre belief needs to be reported by the same person (e.g. Cotards).
- Elaborated vs Circumscribed: Elaborated delusions are ones where the person draws consequences from the delusional state and forms other beliefs that revolve around the theme of the delusion. For example, a person with Capgras develop paranoid thoughts related to the content of the delusion, that the impostor has evil intentions and will cause harm. The story grows and grows – get delusional webs of beliefs. Circumscribed delusions are isolated from the rest of the persons beliefs. Monothematic delusions tend to be circumscribed whereas polythematic delusions tend to be elaborated.

Explanationist Account of Delusions

- Associated with the theories of Brendan Maher. His idea was that some underlying biological problem led to stranger experiences, which people then tried to interpret.
- He thought that delusional belief is a normal response to an abnormal experience. For example, a woman who thought head was full of bees. Seems that may have been having buzzing sound in head caused by some physiological problem.
- There is a lot of evidence that there is an underlying biological problem that is associated with delusions i.e. that are due to brain damage including: stroke, tumour, trauma, dementing illness.
- The account faces a number of problems:
    - 1. There are cases of delusions when there is no evidence of brain damage. Is this because there is no damage, or because we cannot detect the damage?
    - 2. If the problem is biological then need to give an account of the relation between the damage to the brain and the delusional mental states. What relation does brain damage have to thinking that one's wife is an imposter?

- 3. The thinking of delusional people is clearly not normal. There seems to be a big leap from hearing buzzing to thinking your head is full of bees. A more obvious and sensible conclusion would be that one has some form of brain damage.

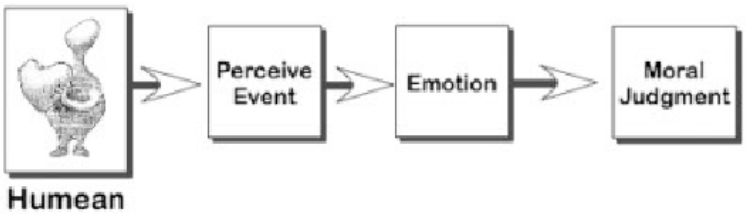
Multi-factor Accounts of Delusions

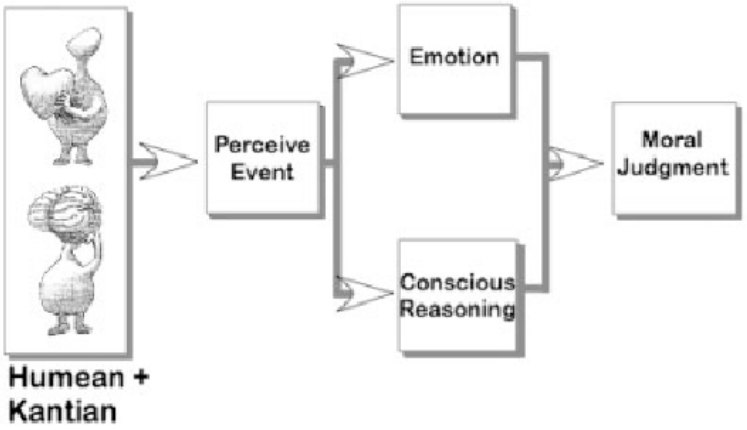
- One-factor theorists hold that anomolous experiences are enough to explain a delusion.
- Two-factor theorists hold that the unusual experience is only one factor in the formation of the delusion.
- One of the strongest arguments for a two factor theory is that there seem to be people who have anomalous experiences like delusional people, but who do not endorse the delusional beliefs. In the case of schizophrenia this can happen in the same person at different times.
- The reasoning in delusion may be merely a reasoning bias that all people (normal and delusional) suffer from, or it might result from a deficit, that is it might be an abnormal type of reasoning.
- Attributional style: we all have a tendency to attribute causes to either internal or external forces. Maybe Cotards subjects have an internal attribution bias whereas Capgras have an external attribution bias. Problem with this is that some people have both delusions.
- Conservation vs observational adequacy: Our beliefs seem to be linked together into a web of interconnected ideas. Our beliefs don't exist in isolation. We are conservative when it comes to changing our beliefs, particularly our central ones. But our beliefs also need to fit with what we observe, so there is a tension between revising our beliefs on the basis of new observations and not disturbing our central beliefs. Maybe people with delusions have particular trouble with belief revision.
- Probabilistic reasoning bias: Perhaps people with delusions don't take into account the probability of various events and instead just jump to conclusions. Problem is that no human seems to be very good at probabilistic reasoning.

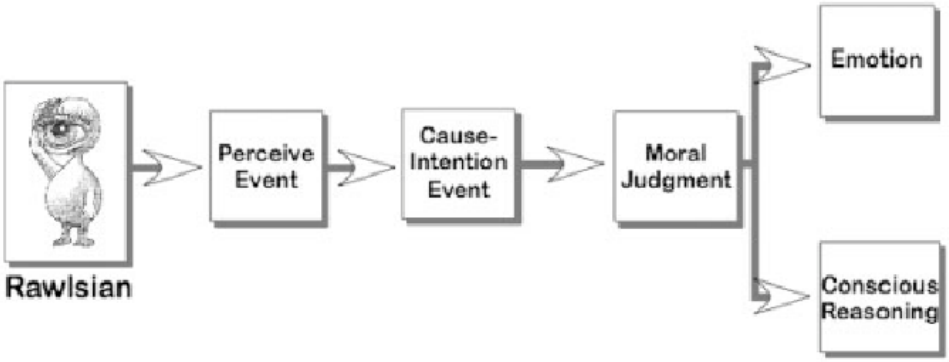
Topic 5: Morality and Moral Grammar

Philosophical Theories of Moral Judgement

- Simple Sentimentalism: Sentimentalists think that moral judgements are essentially expressions of emotional responses to situations, acts, or agents. Jesse Prinz says that 'When we utter "ought" we express our own sentiments and factual knowledge is not sufficient for having sentiments'.
- Externalism: Claims about what is morally required are to be thought of as judgments about the correct application of a set of moral standards to some situation, arrived at through explicit procedural reasoning. Moral judgments are argued to be analogous to judgements about the requirements of etiquette or the law. Our beliefs in these domains have no necessary connection to motivation.
- Internalism: Holds that there is an internal connection between moral judgment and motivation to act. Moral judgments are distinguished from judgments of mere liking and disliking by being answerable to reasons. Genuine moral judgments must be reason-responsive.







The Role of Cognitive Science

- Evidence from the sciences may reshape philosophical debates by e.g., undermining traditional philosophical distinctions, between affect and reason.
- We might want to impose a minimal psychological realism constraint on our meta-ethical theories.
- One idea along these lines is that the cognitive penetrability of moral judgments seems critical if internalist criteria for moral agency are to be met.

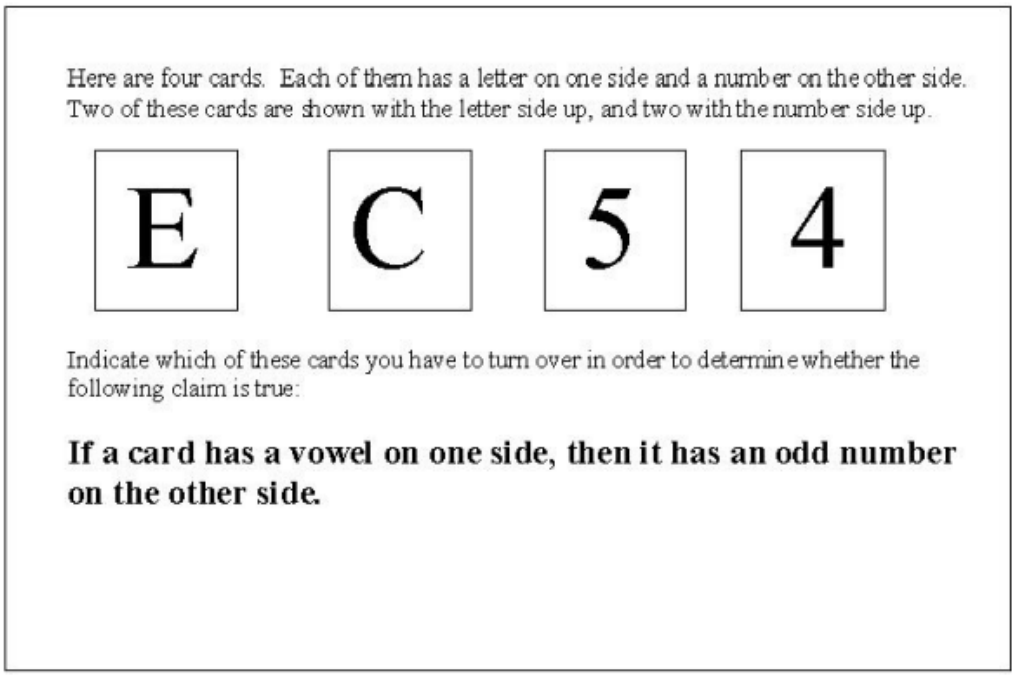
Hauser's Moral Grammar

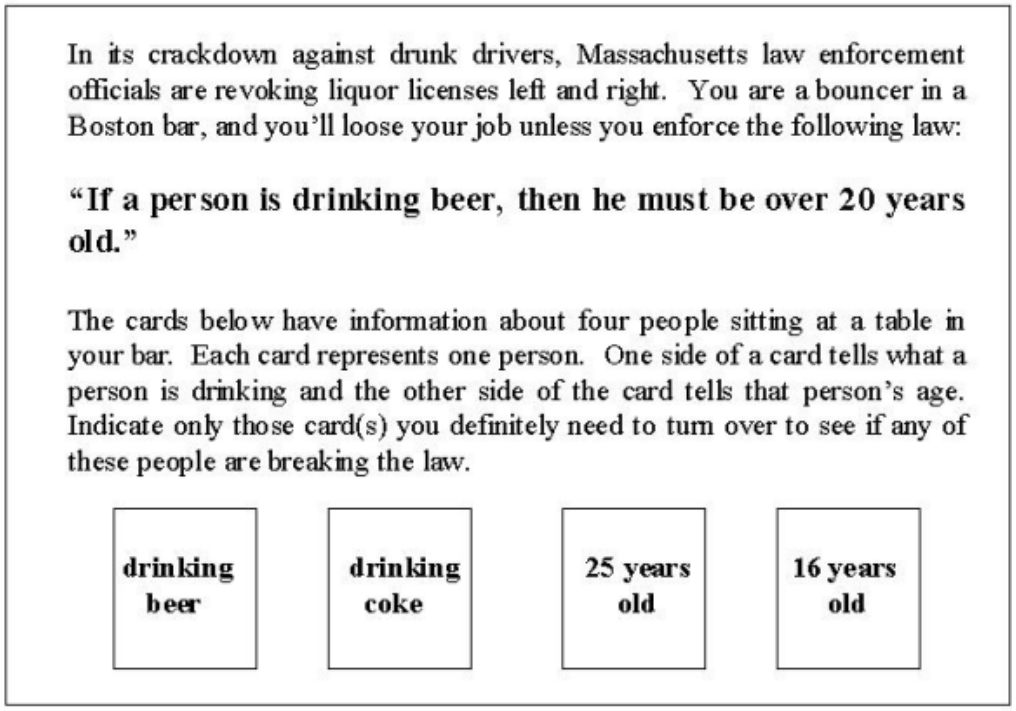
- Hauser argues that we have a universal moral grammar in the same way that we have a universal linguistic grammar.
- Universal grammar: the set of all innate principles available to each child for acquiring a natural language.
- We are generally not consciously aware of these rules and principles, even if we have strong operative knowledge as to how to deploy them. Hauser believes that this is similarly the case for moral judgements.

- Hauser's claim: "We are endowed with a moral faculty that operates over the causal-intentional properties of actions and events as they connect to particular consequences. On this model, we are endowed with a moral faculty that generates judgments about permissible and forbidden actions prior to the involvement of our emotions and systems of conscious, rational deliberation. We posit a theory of universal moral grammar which consists of the principles and parameters that are part and parcel of this biological endowment."
- Hauser argues that as with grammar, moral principles are operative but not expressed, and "are only discoverable with the tools of science ...the operative principles of the moral faculty may do all the heavy lifting, generating a moral verdict that may or may not generate an emotion or a process of rational and principled deliberation."
- This view holds that the unconscious appraisal mechanism provides an analysis of causes and consequences of actions, and represents moral competence and is responsible for judgment. Emotions are only related to moral performance (whether we act in accordance with our moral judgment).
- Rawls similarly thought that the principles underlying our moral intuitions might be unconscious and inaccessible: "There is no reason to assume that our sense of justice can be adequately characterised by familiar commonsense precepts, or derived from more obvious learning principles. A correct account of moral capacities will certainly involve theoretical constructs and principles which go beyond the norms and standards cited in everyday life."

Evidence for a Moral Grammar

- Derives largely from research on responses by ordinary people to hypothetical moral dilemmas, including the trolley problem and the online Moral Sense Test.
- In terms of interpreting the results, it seems that neither pure deontological rules (don't kill) nor utilitarian considerations (maximise utility) account for the intuitions generated.
- Results were consistent across groups yet only 30% provided "sufficient" justifications.
- Can we detect principles that guide the participant's responses? Are these principles accessible to the participants?
- A principle which explains a good deal of the variation in responses to the cases is a prohibition on intentionally using someone as a mere means to a positive outcome. Most subjects could not explicitly access this justification and it's plausible to think that most people are never explicitly taught it.
- Hauser et al (2008) think the results support the notions that "some forms of moral judgment are universal and mediated by unconscious and inaccessible principles".
- The Wason Selection Task is also advanced as evidence for a moral module. In this case few people can correctly answer an abstract decision problem, but most people find it easy to get the right answer if it is applied to a specific case of violating a moral/social norm (teenage drinking rules). It has been argued that this suggests that when people reason about deontic matters they are probably not employing domain general procedures.
- Based on this evidence proposed principles of this moral grammar include:
    - Doctrine of double effect (intention-foresight distinction)
    - Acts-omissions distinction
    - Cheater detection





Arguments Against a Moral Grammar

- Shaun Nichols suggests that we have both deontological and utilitarian moral rules or systems, and that these systems operate partly independently. Thus when we assess the morality of a situation, it must pass both deontological and utilitarian considerations to be judged as morally acceptable. This two-system model would provide an internal explanation for why the doctrine of double effect is externally adequate to our intuitions. However, the two-system model does not require that the principle itself is internally represented at all.
- Sterelny's Challenge: "We have both tacit and explicit moral cognition, and the systems interact. Yet none of the moral grammarians have a model of the interaction between the

fast, tacit, automatic system and the slow conscious system." This is a problem for an account like Hauser's, since it would require us to think that moral reasoning is epiphenomenal - it has no causal role. We only have specific moral responses to specific situations, and we could not account for moral generalisations at all.

- Haidt thinks that moral reasoning *is* epiphenomenal. Moral intuitions are the result of automatic, unconscious processes like aesthetic judgements, and if moral reasoning occurs it is generally in social contexts where we wish to influence other people, or justify something to ourselves.
- Kennett&Fine (2009) argue that there is a complex interplay between automatic and deliberative mental processes in moral judgment formation. The particular value judgments we make in line with our reflective views could become fast and habitual over time, just like learning to drive and other skilled activities - no surprise that judgments about straightforward situations are fast. The capacity to deliberate upon and regulate one's automatic affective responses seems essential to moral agency. We can also learn to override automatic responses; not always a case of justifying them post-hoc.

Frontal Lobe Damage

- In the case of Phineas Gage and other patients who have experience lesions to the frontal lobe, it is typical to display anti-social behaviours. These appear to be linked to difficulties with emotional regulation.
- Gage for example: "Seems to have lost his ability to observe previously acquired social conventions and ethical rules as well as ability to anticipate and plan for the future. But general intellectual capabilities were intact."
- This seems to support Nichols' view: once the rules are decoupled from emotion they lose their normative force.

Mindreading

- Evidence for the importance of only minimal relevance of mindreading is that double dissociation between deficits of theory of mind (autism) and anti-social behaviours (psychopathy).
- Blair (1995) has proposed that the psychopaths' poor performance on the moral/conventional task arises from impairment in processing fear and distress cues, i.e., impairment in the recognition of a victim.
- One psychopath: When asked how he would feel in such a situation (bank robbery), his reply made no reference to bodily sensations. He said such things as, "I'd give you the money"; "I'd think of ways to get the drop on you"; "I'd try to get my ass out of there". When asked how he would feel, not what he would think or do, he seemed perplexed. (Hare, 1993, pp. 53-54).
- Psychopaths are inevitably deficient in moral understanding. By contrast autistic children do show responsiveness/arousal to others distress and do make a moral/conventional distinction. Unlike psychopaths some autistic people display high levels of moral conscientiousness (Kennett 2002).

Topic 6: Interpreting Neuroimaging Research

Problems with Neuroimaging

- Functional Magnetic Resonance Imaging doesn't measure neural activity. Looks at blood flow, but even does this indirectly
    - Compare blood flow during experimental and control task
    - Perform statistical tests for differences in blood flow in different regions
    - Colour blobs represent statistical significance, not activation
- Relies on successful isolation of the two mental states of interest
    - Relatively easy for basic sensory processing
    - Very challenging for more complex mental concepts
- Nuisance variables (e.g., caffeine intake, breathing rate, menstrual cycle) create noise
    - Exacerbated by small sample sizes (Wallentin 2009, Brain & Language 108, 175)
    - Unfortunately because of expense imaging studies often have small sample sizes
- Healthy controversies over appropriate statistical techniques
    - Standard thresholds for significance yield 'significant' activations in a dead salmon (Bennett et al 2009, NeuroImage 47, S125)
    - So they claim we need to tighten up thresholds of statistical significance
    - 'Voodoo' correlations (Vul et al 2009 Perspectives Psych Sci 4, 274)
- These problems often result in very inconsistent outcomes across studies, for example the alleged finding of sex differences in language lateralisation.

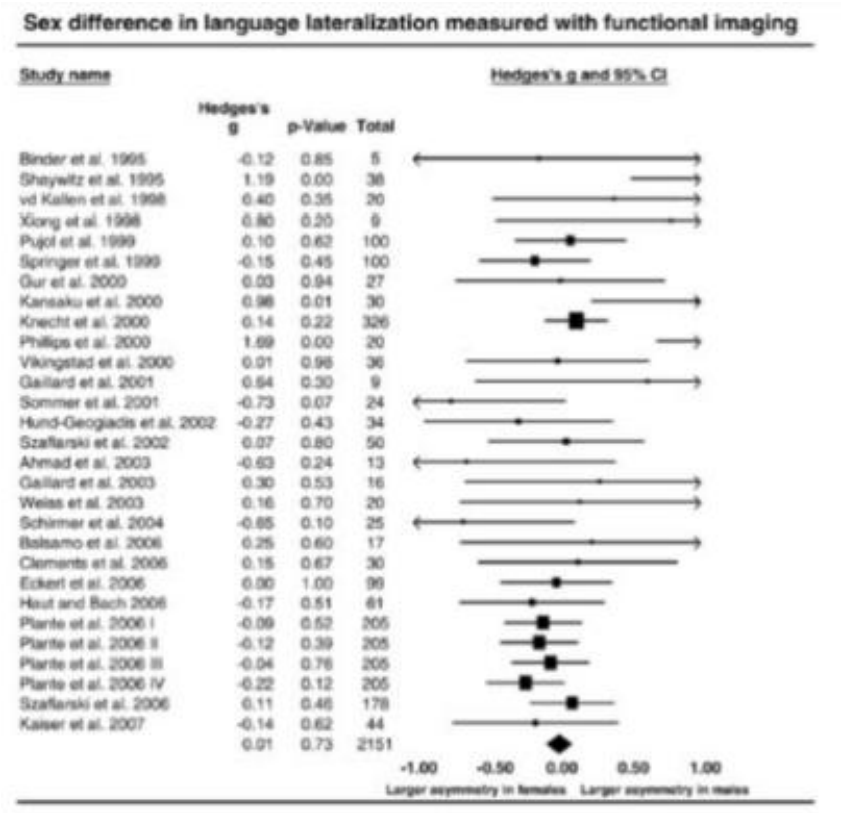
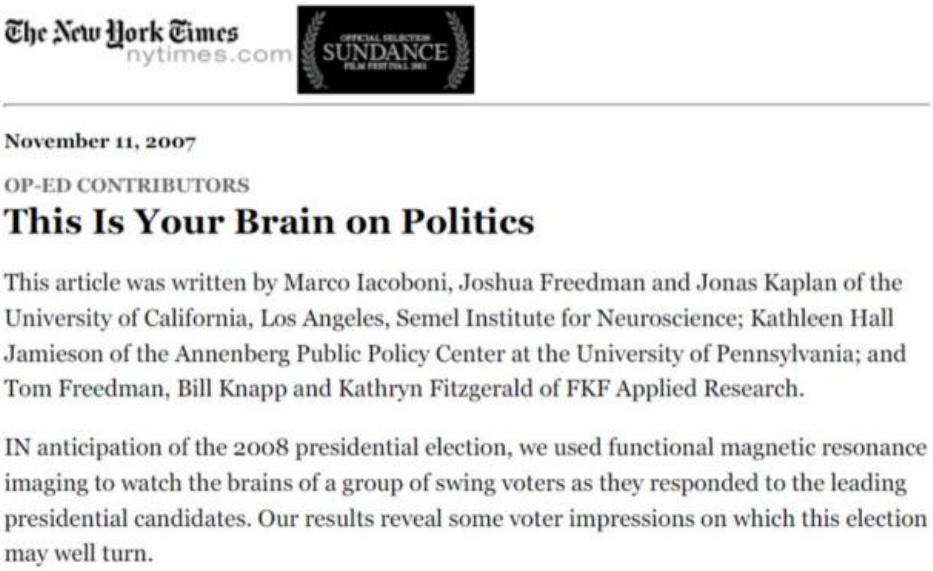


Fig. 4 – Sex differences in asymmetry of language activation.

Interpretation of Results

- Forward inference:
  "Brain region X was active while participants performed cognitive operation Y" ->
  "Brain region X is therefore involved in cognitive operation Y"
- But there are problems with this reasoning, as just because a brain region is active, does not mean that region is either necessary or sufficient for engagement of that mental process.
- For example, many neuroimaging studies had implicated the medial prefrontal cortex in theory of mind tasks, but Bird et al (2004, Brain 127, 914) patient with a medial prefrontal lesion performed normally on a theory of mind test.
- Reverse inference:
  "Brain region X was activated, therefore mental process Y was taking place in X"
- This reasoning is quite fallible, as it will only be valid of *only* mental process Y activates brain region X, but most brain regions activated by a variety of mental processes.
- For example, the anterior cingulate gyrus is active in so many studies that one neuroscientist calls it the "on-button".
- 'More activity' does not necessarily mean 'psychologically more signal'; it might not mean that mind is working harder or that the person is better, since as people improve sometimes activity in area decreases.
- 'No activation' does not mean no activation, since what the scans measure is the *difference* in signal between control and experimental tasks. So an area might not "light up" not because it is not working but because it is involved in both tasks.
- Temporal resolution is still only on the order of 1 second, which is longer than many cognitive tasks.
- Example of poor interpretation of neuroimaging results:



The New York Times nytimes.com

November 11, 2007

OP-ED CONTRIBUTORS

This Is Your Brain on Politics

This article was written by Marco Iacoboni, Joshua Freedman and Jonas Kaplan of the University of California, Los Angeles, Semel Institute for Neuroscience; Kathleen Hall Jamieson of the Annenberg Public Policy Center at the University of Pennsylvania; and Tom Freedman, Bill Knapp and Kathryn Fitzgerald of FKF Applied Research.

IN anticipation of the 2008 presidential election, we used functional magnetic resonance imaging to watch the brains of a group of swing voters as they responded to the leading presidential candidates. Our results reveal some voter impressions on which this election may well turn.

- There is no straight-forward connection between brain structure and brain function. For example, greater male lateralisation of language has been argued to explain women's superior abilities in language, mulit-tasking and empathising, because have more parts of the brain doing these things. However, there is really no reason to think that greater connectivity of brain regions translates to greater connectivity of thoughts.

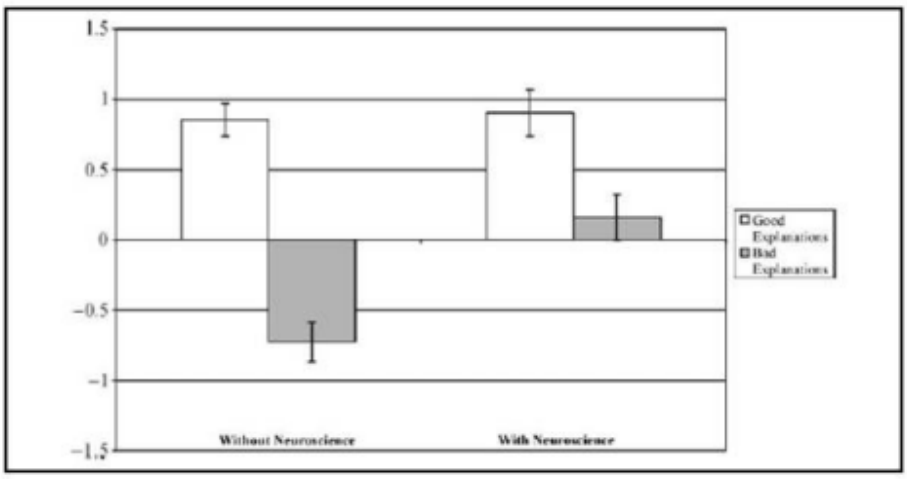
- A skew towards local connectivity is “compatible with strong systemizing, because systemizing involves a narrow attentional focus to local information, in order to understand each part of a system.” Baron-Cohen et al. (2005), Science, 310, 819
- “Better interhemispheric connectivity” of the female brain “confers an advantage in language” while “males may excel in tasks requiring focal activation of the visual association cortex.” Halpern et al. (2007), Psych Science Public Interest, 8, 1
- Response: “May as well say hairier body so fuzzier thinker. Or that human beings are capable of fixing fuses because the brain uses electricity.” (Ian Gold, pers. com.)

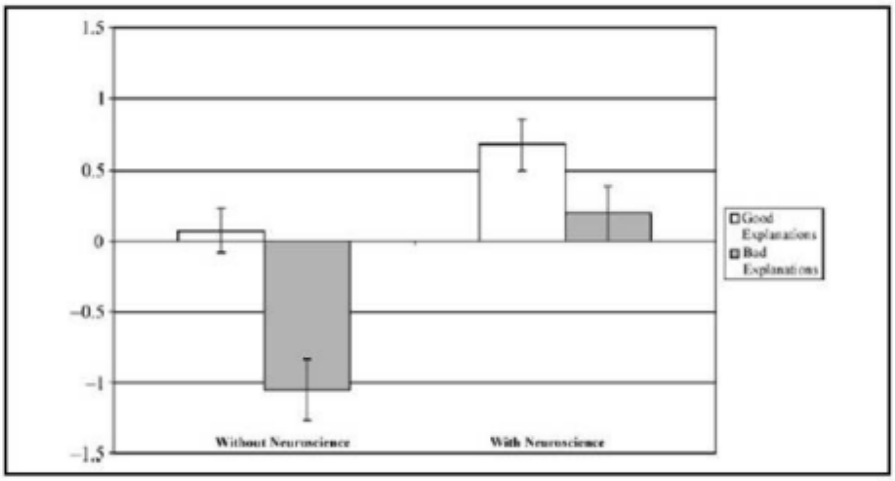
Explanatory Power

- Neurorealism: evidence from experimental studies and from scientific practise seems to show that we may view neuroimaging research as being "more scientific" or objective than behavioural techniques.
- Racine et al (2010) Social Science & Medicine 71, 725: “Neuroimaging research yields direct data on brain function despite the complexities of data acquisition and image processing involved. Observed brain activation patterns are, as a result, portrayed as the ultimate proof that a phenomenon is real, objective, and effective.”
- Examples from Racine et al (2010) Social Science & Medicine 71, 725:
  - “Now scientists say the feeling is not only real, but they can show what happens in the brain to cause it.
  - “A relatively new form of brain imaging provides visual proof that acupuncture alleviates pain.”
  - “Fat really does bring pleasure.”
- Example from the scientific literature Steinberg & Scott 2003, Am Psychol, 58, 1009:
  - “Most of the developmental research on cognitive and psychosocial functioning in adolescence measures behaviors, self-perceptions, or attitudes, but mounting evidence suggests that at least some of the differences between adults and adolescents have neuropsychological and neurobiological underpinnings.”
  - But all behaviours have a neural underpinning!
- Even if we establish brain causation, what does it mean? Finding a difference in the brain doesn’t by itself say anything about why this difference exists or what caused it, or what difference it makes in psychological processing.
- “The causation of virtually any complex behavior is affected by psychological and sociological variables, even when brain causation has been identified… Monolithic brain explanation of complex behavior is almost always radically incomplete.” Morse 2005

Table 1. Sample Item

	Good Explanation	*Bad Explanation*
Without Neuroscience	The researchers claim that this “curse” happens because subjects have trouble switching their point of view to consider what someone else might know, mistakenly projecting their own knowledge onto others.	The researchers claim that this “curse” happens because subjects make more mistakes when they have to judge the knowledge of others. People are much better at judging what they themselves know.
With Neuroscience	**Brain scans indicate** that this “curse” happens because **of the frontal lobe brain circuitry known to be involved in self-knowledge.** Subjects have trouble switching their point of view to consider what someone else might know, mistakenly projecting their own knowledge onto others.	**Brain scans indicate** that this “curse” happens because **of the frontal lobe brain circuitry known to be involved in self-knowledge.** Subjects make more mistakes when they have to judge the knowledge of others. People are much better at judging what they themselves know.





Topic 7: Neurolaw and Free Will

What is Neurolaw?

- Neurolaw is a new area of research, study and legal practice at the intersection of three disciplines: Law, Mind Sciences, and Philosophy.
- The new mind sciences challenge some of our current legal practices
    - Some researchers argue that these modern sciences show that some of our current legal practices are based on false assumptions
    - For instance, they argue that science shows that responsibility does not exist (e.g. see Greene & Cogen 2004)
    - In the case of Greene and Cohen, they insist among other things that we should stop punishing people for retributive reasons.
- Philosophy provides the conceptual tools necessary to translate the ideas, problems and solutions in one field (science) to the ideas, problems and solutions in the other field (law).
- For instance, philosophical compatibilists like Fischer and Ravizza (1998) have developed a theory of responsibility according to which the above-mentioned scientific findings need not necessarily be interpreted as posing a challenge to the notion of responsibility.

Neuro-Technologies and Neurotechniques

	diagnostic *observation/reading*		**intervention** *alteration/writing*
psychopharmacology	blood chemistry analysis		psychotropic drugs
invasive techniques (*intra*cranial)	ECoG *ElectroCorticoGram*		DBS *Deep Brain Stimulation* brain surgery
non-invasive techniques (*trans*cranial)	EEG *ElectroEncephaloGram* MEG *MagnetoEncephaloGram* NIRS *Near Infra-Red Spectroscopy* TCDS *TransCranial Doppler Spectroscopy*		ECT *ElectroConvulsive Therapy* TMS *Transcranial Magnetic Stimulation*
radiology	structural	functional	
ionizing radiation	CT/CAT *Computed Axial Tomography (e.g. x rays)*		radiation therapy *(e.g. x rays, gamma rays)*
radioactive isotopes		SPECT & PET *Single Photon Emission Computed Tomography and Positron Emission Tomography*	radioisotope therapy
magnetic fields	MRI *Magnetic Resonance Imaging*	fMRI *functional Magnetic Resonance Imaging*	
	DTI *Diffusion Tensor Imaging* DSI *Diffusion Spectral Imaging*		

Deception Detection

- In court cases, witnesses can lie, and even if they do not intend to deceive, memory is highly constructive and thus prone to be unreliable.
- The Polygraph was one attempt to do this. It measures skin conductance, the heart's pulse, and breathing rate. The idea is that when a person attempts to deceive another theirperipheral nervous system will give away signs that they are doing this.
- More modern approaches have built on this idea but takes measurements of the central nervous system. Works on the presupposition that lying/deception is more difficult than telling the truth i.e. that it requires extra effort, and perhaps special parts of the brain or special brain circuits.
- Central idea: When someone who is personally familiar with some aspect of a crime sceneviews a picture of that crime scene, their brain will behave differently to the brain of someone who is not personally familiar with that aspect of the crime scene.
- Two companies have tried to develop and commercialise methods to detect deception and guilty knowledge: Cephos Corporation http://www.cephoscorp.com/ and No Lie MRI http://www.noliemri.com/.
- However the science upon which these techniques and technologies are based has not been peer reviewed, since the companies wish to keep secret how their techniques work.
- The type of "lying" or "deception" used in the experiments is very different from real world lying or deception: in the lab the subject is requested to tell the experimenter something that they know to be untrue, which is not how it works in the real world.

Mind-reading

- Another approach to overcoming the fact that people lie, or just plainly that they don't want to tell us what they know, could be "mind reading", though not in the sense that has so far been discussed in this course, but rather in the literal sense — i.e. looking into people's brains to figure out what they know and to pull it out of their brains.
- There has been some very limited successes in this area: it is possible to use neuroimaging technology to detect whether a person is thinking of a place (e.g. their home) or a face (e.g. their grandmother's)
- But the science is weak because we still do not have enough of an understanding of exactly how knowledge is encoded in the brain.
- The problem is that even at a basic conceptual level once we think more closely about what mind reading might involve, we strike some difficult problems.

Lie Prevention

- Another approach to overcoming the fact that people lie is to develop intervention (rather than diagnostic) techniques and technologies: temporarily disable people's ability to lie and deceive.
- To the extent that lying requires special brain modules or circuits, it is plausible to suppose that once we figure out which brain areas and/or circuits are needed for lying, we might be able to develop techniques to modify how those brain areas function.
- The drug commonly known by its trademark name "Sodium Pentathol" (the so-called "truth drug") is one example of a crude direct brain intervention technique designed to extract the truth from subjects.
- Another approach would be to use a technology like transcranial magnetic stimulation (TMS) to temporarily disable those parts of the brain which are used to deceive.

Crime Prevention

- Perhaps we can find ways to assess the likelihood that someone will commit a crime by examining their brain? Perhaps there might be such a thing as a "criminal mind".
- One problem is that crimes are defined in specific ways in statutory law, and it is unlikely that these complex ideas map very well (if at all) to features of the mind or brain.
- A more promising approach is to examine the correlation between crime and certain pathological mental conditions.
- For example, crime has been correlated with psychopathy, and psychopaths seem to have brains that are discernably different from non-psychopaths' brains (e.g. see Blair 2007).
- Maybe paedophilia is also a condition that can be diagnosed through brain imaging techniques?
- There is also the question of the use of these techniques for sentencing and parole hearings, where many ethical issues also come into play.

Responsibility

- In order to attribute responsibility to someone you need to establish both actus reus and mens rea.

- Actus reus: was a criminal act committed? This means a conscious action, and not just some spontaneous uncontrollable bodily behaviour, like a reflex or spasm. Neuroscience might be able to help in some tricky cases (like sleep walking)
- Mens rea: is the person blameworthy, did they have a guilty mind? If can show that your action was a consequence e.g. of a delusion then that might be a reason not to blame you, or to diminish the blame.
- Example: A school teacher who suddenly develops interest in child pornography and then makes advance on daughter
    - Sent to sex-offenders treatment
    - At course propositions other people so ejected from court
    - Goes to ER with bad headache and has a variety of neurological symptoms
    - Find an enormous cyst that is putting pressure on brain
    - Remove cyst and within 2 weeks all pedophellia symptoms disappear
    - Same thing happens a year later
- Here technology has uncovered a neurological explanation and stopped us from attributing the blame we would otherwise have done.
- What about people who developed people in child pornography much earlier in life? Could they also have an undiscovered neurological condition that would also exonerate them?
- Do we want to treat any brain difference as a disorder? What is a normal brain?
    - Seems OK to do in the case of the teacher, but should we generalise this?
    - Presumably many individual differences have a neurological underpinnings, e.g. for homosexuality
    - Does any significant neurological difference mean that it is abnormal?
    - What are the neurologically differences that ought not to be there from those that are just part of the acceptable differences between people?
    - Our cultural norms change regarding what is a disorder
    - What is the difference between a character flaw (blameworthy) and lacking a capacity or having a disorder (not blameworthy)?
- Key point is that finding a neural difference does not enable us to determine whether something is a disorder versus a difference that is within the normal range.

Free Will and Neuroscience

- Green and Cohen (2004) argue that neuroscience reveals that the brain is just a machine, since what happens is just a chain of physical causal process. They argue this undermines free will and therefore responsibility. Instead of punishing we should predict and treat.
- The problem with this argument is that it is not clear what neuroscience really adds beyond existing (and very old) arguments about determinism.
- The traditional options regarding free will are:
    - Hard determinism
        - Accepts that free will and determinism are incompatible
        - Bites the bullet and claims we don't have free will
    - Libertariansim
        - Accepts that free will and determinism are incompatible
        - Bites the bullet and claims that determinism is false
    - Compatibilism

        - Freedom is compatible with determinism
        - Freedom isn't the ability to have uncaused behavior
        - It is the ability to cause things in the right way, or to have the right capacities
- The idea here is that causation is not an excuse, but lacking the relevant capacities to choose can be.

Libet's Study

- Libet conducted a famous study in which he compared the time of awareness of intention to act with cerebral processes that generate the act (the readiness potential), and found that the cerebral processes preceeded the awareness of the intention.
- This suggests that this means that the volitional process is initiated unconsciously and that unconscious initiation of the voluntary process appears to mean that conscious free will could not actually 'tell' the brain to begin its preparation to carry out a voluntary act.
- The interpretation that Libet gave is that your conscious decision to behave does not cause your behaviour. It comes after your brain had already decided to act, and themselves are causally inefficacious. Instead, unconscious neural processes cause your behaviour. It doesn't seem that you are responsible for your unconscious neural processes.
- One problem with this interpretation is that it is impossible to know that the readiness potential that Libet measures actually corresponds to – is it the cause of our decision or merely the beginning of the process of deciding, or just an urge to act in some way? We don't know.
- Another criticism is that in Libet's study, subjects were instructed to decide spontaneously, and under such unusual circumstances we may expect that the brain activity that precedes action is very different from when we act in accordance with consideration and our reasons.

Biology and Culpability

- Behavioural genetics also raises similar issues. For example, consider the case of people who lack the enzyme monoamine oxidase A, which results in an inability to break down serotonin.
- Excess seratonin leads to violent dispositions if brought up in a violent environment.
- Therefore these people have a much more uphill battle; they are much more easily angered, and this is because of their genes and their upbringing, which they are not at fault for.
- But the fact that we have a causal explanation doesn't excuse someone.
- Instead we need to look at what capacities they do or don't have
- Note that social factors can cause a lack of relevant capacities, e.g. growing up in abusive environment may cause certain incapacities.