A selection of essays in cognitive neuropsychology

Kelly Alexandra Roe

2008

Introduction

Fall 2008 I audited the first of a pair of graduate level courses in cognitive neuropsychology at Duke. What follows are my essays that have been reprinted in a more sensible font.

Contents

Chapter 1

Behaviourism

In the first half of the 20th century behaviourism largely failed as an approach to understanding cognitive functions. What was the root cause of this failure? Are any of the concepts used in cognitive neuroscience today in similar jeopardy?

Behaviourism doesn't assist our understanding of cognitive functions for the most part because it doesn't purport to. Methodological behaviourism is an attempt to reconceptualize psychology as the science of behaviour, rather than a science of mental states or cognitive functions. As such, methodological behaviourists maintain that we should either eliminate terms that refer to mental states from the vocabulary of a science of psychology, or that we should operationalize them so that they refer to observable behaviour - or, more plausibly, to dispositions to behave. Dispositions are thought to be reputable scientific notions because they can be operationalized into observable behaviours. To say that a substance is soluble, for example, is just to say that in certain sorts of circumstances (such as being placed in water) certain sorts of observations (dissolving) would be found. Similarly, to say that a person is in a certain kind of mental state (e.g., pain) is just to say that if they were in certain kinds of circumstances then they would behave in a certain kind of way.

There are observations that the behaviourist paradigm seems incapable of explaining, however. Stimuli, response, and reinforcement history are unable to account for such phenomena as: filial imprinting, critical learning thresholds, fixed action patterns, species specific learning (e.g., easy to teach a hen to peck a key and a rat to push a lever but the converse is not the case), and species specific behaviours such as language (human infants will acquire syntax but not the infants of other species). Advances in ethology, genetics, and neurophysiology thus provided alternative paradigms for theorists and researchers who were interested in explaining those phenomena. Advances in computer science, linguistics, and cognitive psychology also provided alternative paradigms for researchers and theorists who were interested in ex- plaining phenomena that didn't seem to be explainable from within the behaviourist paradigm. Examples of such phenomena include: a computer being programmed to perform an 'intelligent task' despite the lack of learning and reinforcement history, the seemingly rule governed combinatorial structure shared by syntax, logical thought, and computation, and limitations on what could be learned or remembered.

The first movement away from behaviourism thus came from the idea that there must be a categorical basis to dispositions. If we return to solubility it seems that there is something about the internal nature of substance in virtue of which it has its dispositional profile. To learn about the nature of bonds between atoms would seem to further our knowledge by explaining why the substance has the disposition to dissolve. Analogously, it was argued that behavioural dispositions must be realized in neurophysiology and that getting clearer on the categorical basis of dispositions was as legitimate a branch of psychology as getting clearer on the laws governing dispositional profiles. Thus it was argued that solubility should be identified with the categorical chemical bonds and mental states should be identified with the categorical neurophysiological underpinnings. The second movement away from behaviourism came from the notion that an adequate dispositional anal-

ysis required a more complex structure than that posited by behaviourists. While behaviourists limited themselves for the most part to explaining relatively simple stimulus - response (S-R) pairs, complex behavioural dispositions were best explained by positing intervening layers that interacted with other intervening layers (e.g., S - R/S - R/S - R). Getting clearer on the computational or functional profile of mental processing thus seemed to be as legitimate a part of psychology as the laws governing stimulus - response interactions. Thus it was argued that mental states should be defined according to their informational or computational structure in a network rather than with dispositions to behave.

Despite popular opinion that the behaviourist paradigm is dead and buried, behaviourism continues to be a profitable paradigm in psychology. This is evidenced by the number of journals devoted to it, the number of researchers and theorists who identify with it, and the considerable research grants that continue to be given to behaviourists investigating such things as: the management and treatment of autism, addiction, and intellectual handicap, the discriminations that animals are able to make, how hard animals are prepared to work for various commodities (e.g., access to scratching litter). We have seen how there are limitations to the observations that the paradigm was suitable to explain, however. The development of alternative paradigms resulted in theorists and researchers reorienting their attention towards explaining phenomena that were anomalous for the behaviourist framework. Thus what once was the predominant paradigm for psychology came to be but one of many.

Cognitive neuropsychologists often claim that cognitive neuropsychology provides a paradigm that successfully integrates the relatively high level functional, informational, or representational mechanisms[1] of cognitive psychology and the relatively low level neurophysiological mechanisms of persons.

---

[1]'Mechanism' should be understood as a shorthand for 'state', 'process', 'entity' etc.

One proceeds by 'popping the hood' (looking inside the box at the neurophysiology) which is thought to provide evidence to support or dis-confirm different cognitive psychological theories of the mechanisms involved in mental processing. It is important to note that in order for neurophysiological mechanisms to be relevant for supporting or falsifying theories about cognitive psychological mechanisms they must (in the first instance) be found to be correlated, however. Insofar as computationally or representationally defined mechanisms fail to correlate[2] with physiologically defined mechanisms there is no integrated science of cognitive neuropsychology.

The point of the software / hardware distinction isn't to deny that software (informational or computational) mechanisms must be realized or implemented on some (neurophysiological or silicone) hardware mechanisms or other. Rather, the point of the distinction is to show that what counts as a kind of mechanism at the software level (e.g., a 'save function' or a 'pain') may be massively multiply realized at the hardware level (e.g., neurological differences within an individual over time, between different individuals, between different species, between persons and computers...). Such massive multiple realizability is thought to undermine the plausibility of the lower level science attempting to identify mental mechanisms with kinds of hardware mechanisms (to provide the categorical basis of the suitably enriched behavioural disposition). Insofar as the kinds of mechanisms posited by cognitive psychology fail to be correlated with the kinds of mechanisms posited by neurophysiology it is hard to see how findings pertaining to one can be used to either support or falsify hypotheses pertaining to the other. In the absence of one-one type mappings neurophysiology seems as irrelevant for explaining cognitive mechanisms and functions as particle physics is irrelevant for explaining laws in economics such as Fisher's law. Notice that this is not to say that Fisher's law could be instantiated without there being some distribution of physical particles or other. It is simply to deny that

---

[2]One way to defend the box factory / solubility analogy would be to maintain that there are correlations to be had after all. I'd be interested in seeing them.

the language (and ontology of mechanisms) of particle physics is relevant for capturing the counter-factuals (enriched behavioural dispositions) that are supported from the language (and ontology of mechanisms) in economics.

The failure of correlation suggests two broadly different ways that the future of psychology could go (with there being two options within each broadly different approach):

- 1a: We should revise the mechanisms, entities, and processes posited by cognitive psychology such that they do correlate with neurophysiological states (e.g., by stipulating that they need to be operationalized into talk of kinds of neurophysiological mechanisms).

- 1b: Talk of cognitive psychological mechanisms should be eliminated from a science of psychology.

- 2a: We should revise the mechanisms, entities, and processes posited by neurophysiologists so as to ensure correlation (e.g., by stipulating that kinds of neurophysiological states need to be operationalized so that they are functionally or computationally defined).

- 2b: Talk of neurophysiological mechanisms should be eliminated from a science of psychology.

Critics maintain that cognitive neuroscience recommends that we reconceptualize the science of the mind/brain as one in which which mental mechanisms, processes, and structures, are either operationalized into talk of neurological states in a way that they are not in the science of cognitive psychology - or that they simply be eliminated from a science of psychology. While cognitive neuroscientists and neurophysiologists often tend to recommend the elimination of what doesn't seem profitable for the level of analysis that they are interested in the fate of behaviourism seems to warn that redefining the subject matter won't give lasting success as theorists become interested in answering questions (e.g., about functionally defined computational states) for which the paradigm is ill suited.

Chapter 2

Perception

How would you describe the biological purpose of sensory perception?

One intuition that we have about sensory perception is known as the 'transparency of experience'[1]. The intuition is that our perceptual experience tracks, co-varies with, or represents certain magnitudes or values of distal objective properties in the world[2]. If we consider the process of visual perception then the transparency of experience intuition starts to seem problematic, however. We have objective magnitudes of distal objective properties such as reflectance, illumination, and transmittance (to name just a few). The value or magnitude of these properties results in a relatively proximal objective luminance magnitude which is the amount of light that reaches the retina. The luminance magnitude results in a sensory perceptual experience of brightness (where subjects can report on which of several percepts seems brightest).

---

[1] In what follows I shall focus exclusively on visual perceptions even though it might well be the case that a similar story could be told for the percepts of other sensory modalities. I'll also only focus on one aspect of visual perception.

[2] I am concerned about my terminology and I'm not sure what the correct terminology is. The notion that I'm trying to convey is that there are different mind-independent properties (e.g., temperature) that have different magnitudes or values (e.g., 60 degrees or 70 degrees or 80 degrees etc).

The transparency of experience intuition is that perceived brightness tracks, co-varies with, or represents a magnitude of reflectance. A problem is that the magnitude of luminance (from which the brain arrives at a sensory experience) underdetermines the magnitude of the variables or properties that produced it[3]. For example, if we know that (for simplicity) the magnitudes of two properties contribute toward a magnitude of 10, then the magnitude of both of the properties that produced it is ambiguous. The magnitudes could be: 1,9 or 2,8, or 3,7, or 4,6 etc[4]. It seems hard to explain how the brain produces a determinate experience of brightness that is able to track the objective magnitude of reflectance on the basis of a magnitude of luminance when there are many values of reflectance that are consistent with the magnitude of luminance. Given this the question becomes: How is the transparency of experience possible?[5]

One way of attempting to answer this question would be to maintain that different distal magnitude complexes (e.g., 1,9 or 2,8 or 2,7 etc) occur at different relative frequencies in the world[6] If this is so then it might be that the brain represents the distal magnitude complex that most frequently results in the luminance magnitude. This answer seems to leave it as something of a mystery as to how the brain is able to detect the relative frequency of different distal magnitude complexes for a given luminance magnitude given

---

[3]I'm unclear on whether the relationship between luminance magnitude and perceived brightness is one of correlation or whether there are problems with this link, too. I would guess that top down influences would also come to bear.

[4]I suspect that I'm ignoring complicated issues to do with how the magnitudes of variables sum - or otherwise mathematically relate - for a luminance magnitude.

[5]I'd be interested to know whether neurones have been found that are tuned to (at least some aspect of) reflectance and if so whether they are correlated with subjective reports of brightness / some critter pushing a lever or something to signify subjective reports of brightness if that is possible to train. I apologize if the answers to these are out there already... Or ill formed or something... I'm finding this material fairly dense and quite hard to get my head around in such a short space of time.

[6]Worrying about my terminology again, I hope the example helps make my intended meaning clear.

that the luminance magnitude underdetermines the distal magnitude complex, however.

One attempt to explain the mystery is to maintain that instead of regarding sensory percepts to be end products of input mechanisms we should regard them to be the product of past input mechanisms and resulting behavioural responses that came to be selectively reinforced in virtue of the relative frequency of distal magnitude complexes. This would make sense of the intuition that perception must be useful to guide the action of the organism or such a costly system would never have evolved. If it is the case that an objective magnitude complex (1,9, for example) is more frequently the cause of the luminance value than other objective magnitude complexes; And it is the case that a kind of behavioural response has a good outcome for 1,9 but not for other values; And if it is the case that other kinds of behavioural response (that might have good outcomes for low frequency complexes) have bad outcomes when they are produced in response to 1,9; Then given the reinforcement history luminance values could come to be tightly coupled with behavioural responses that are adaptive responses to the most frequently encountered stimulus. This could result in the brain having a bias for a particular magnitude complex that is the most frequent cause of the luminance magnitude[7]

This account provides something of an explanation as to how it is possible that we track the magnitude complexes that are most frequently responsible for luminance values without leaving it as a mystery as to how the brain is able to track (or partly track) the objective frequency of different magnitude complexes that are responsible for the luminance values. It also provides

---

[7]though if the cost-benefit analysis was different then things could get a little messed up e.g., if there was scope for a 'good enough' response to the 1,9 that could become coupled if the 'optimal' response to 1,9 were costly enough when emitted to the slightly less frequent 2,8. I'm not quite sure that this paragraph is very accurate... But I guess the intuitive idea is that natural selection often enough results in 'good enough but not perfect' type solutions especially when improvements are costly.

something of an explanation as to why we might have this capacity - because this is a useful way of an organism guiding its action.

There are a number of assumptions that are required in this explanation, however. It is unclear to me just how plausible each of the assumptions are. One concern is how sensitive such models are to assumptions that are being made (e.g., assigning relative frequencies to distal magnitude complexes, assigning cost and benefit values to responses. It is also unclear to me what features of the world are adaptive for us to track (is that to be determined by tuning?), and that all sensory perceptual content is a product of what once was tightly coupled distal stimulus - behavioural response mechanisms. It is also unclear to me what the role of sensory percepts is (I think the sensory percept was equates with the subjective experience of brightness) and whether the conscious experience of brightness has much to do with action guidance at all (see Libet, for example, and the numerous examples that we have of our action being guided by unconscious visual perceptions that - I'm thinking - don't count as sensory percepts at all?).

Chapter 3

Executive control

One of the core challenges of cognitive neuroscience is translating research findings into information that can improve real- world problems. Choose an applied problem of executive control - e.g., something to do with decision making, impulsiveness, inhibition, etc. - that is manifest in the behaviour of neurologically normal individuals. How could neuroscience data be used to help solve that real-world problem? Be specific about what neuroscience can contribute, above and beyond what can be understood from purely behavioural means.

The flip side of the above mentioned problem would be that of justifying that cognitive neuroscience research can provide data (that is over and above the data available from alternative e.g., behavioural paradigms) that can help improve real-world problems for the purposes of obtaining research funding[1] I am sceptical about the prospects for cognitive neuroscience data assisting in a solution to problems manifest in the behaviour of neurologically normal individuals. I am less sceptical about the prospects for cognitive neuroscience data assisting in a solution to problems manifest in the behaviour of neurologically abnormal individuals, however. In what follows I'll consider what

---

[1]Cognitive neuroscience is not alone - you should try being a philosopher!

10

would have to be the case in order for neuroscientific data to contribute towards economic theories of decision making (with respect to modelling what people actually do). This will involve our considering how to differentiate neuroscientific from behavioural data and interventions, in order for us to see what neuroscience may contribute that is over and above the purely behavioural. I'll end by attempting to justify my claim that it might be more profitable to focus on real world applications with respect to neurologically abnormal individuals[2].

What would have to be the case?

One way in which neuroscience is thought to assist is by providing biological constraints for economic models. In order for it to assist several things must be the case, however: Firstly, the biological constraints need to be neuroscientific rather than being available from different methodological approaches. Secondly, the addition of neuroscientific constraints must result in an economic model with greater generalization power / predictive utility than models that are lacking those constraints (such that the neuroscientific constraints are relevant to the modelling of phenomena that economists are interested in predicting).

It is important to note that the movement from objective probability to utility to subjective probability and subjective utility was motivated from the desire to make better sense of the *behavioural* data and neuroscientific data wasn't the driver for this move. Some have argued that one example of neuroscientific data being used to provide evidence that can either support or falsify economic models is in the case of subjective probability, risk, and ambiguity. We have two broadly different economic models: One-variable models treat subjective probability as a single variable. Two-variable models

---

[2]While the question stipulated 'neurologically normal' I simply don't see better prospects for real-world applications.

treat subjective probability as being composed of two other variables that are worth distinguishing - risk and ambiguity. It has been found that the pattern of brain activation involved in risk and ambiguity are different. One might then maintain that: '[i]f clear differences were to emerge between the patterns of brain activation evoked by risk and by ambiguity - especially within regions critical for decision making - then one could assume that the process of decision making differs for these conditions'. One might then conclude that the neuroscientific data has provided some evidence in support of the two-variable model over the one-variable model.

While patterns of neural activation are clearly relevant for explaining the brain it is less clear that they are relevant for explaining the behaviour that is of interest to economists, however. If two-variable models fare better with their predictions and generalizations then it would seem that economists should adopt a two-variable model given what they are trying to do. If one-variable models fare better with their predictions and generalizations then it would seem that economists should adopt a one-variable model given what they are trying to do. Economic theories seem to be answerable to *behavioural data* and it is hard to see how neurological data is relevant for economic models (though obviously intrinsically interesting in its own right and very useful if we are interested in understanding how brains produce behaviour).

A one-variable economist could maintain that the neurological finding shows subjective probabilities to be multiply realized or distributed in the brain. A two-variable economist could maintain that the neurological finding shows subjective probabilities to consist in two different processes. Conversely, if it was found that there was one neurological process involved in both risk and ambiguity that doesn't support one- variable theories over two-variable theories as two variable theorists could maintain that different tasks can be trained on the same neural network. Since the neurological finding is consistent with both one and two variable models in economics (and since the

converse neurological finding would similarly be consistent with both one and two variable models in economics) neurological findings don't seem to support one economic theory over the other[3].

Differentiating kinds of data and interventions

Differentiating the paradigms (e.g., behaviourist, cognitive, cognitive neuroscientific) is problematic as we shall see, but it is important to begin with a fairly intuitive distinction in order to see what neuroscience may be able to contribute that is over and above the alternative paradigms. While behaviourists concerned themselves with S-R relationships (and how reinforcement can alter that relationship) cognitive psychologists concern themselves with developing models that are answerable to a special class of behavioural data - response time, accuracy of responding, double dissociations e.t.c. What cognitive neuroscience contributes over and above these other two theoretical paradigms is data about neurones and populations of neurones. The contribution of neuroscience would seem to be how we can alter stimuli in order to alter neuron activity and how that altered neural activity can in turn alter behavioural output when generalizations from stimuli to neuronal activity and neuronal activity to behavioural output couldn't be predicted from S-R alone.

---

[3]Note: This is not to say that neurology doesn't constrain the behaviour of individuals. It is clear that particle physics also constrains the behaviour of individuals, however. What is at issue here is whether neurology is more relevant to economic models of what people will do than particle physics.

Motivating Looking to Neurological Abnormality.

It thus seems that neuroscientific data is going to be tied to (fairly direct) neuroscientific interventions and manipulations. Using cognitive neuro scientific (rather than S-R) data to alter peoples behavioural output would seem to involve a fairly direct manipulation of neurology. This is because altering the stimulus would count as a *behavioural* rather than *neurological* intervention - but here we are explicitly interested in what cognitive neuroscience can contribute that is *over and above* that which can be done by behavioural methods alone[4]. There are controversies enough about altering the neurology of neurologically *abnormal* subjects[5]. Advocating the alteration of the neurol ogy of neurologically *normal* subjects would seem to be ethically deeply problematic - and yet it is hard to see how cognitive neuroscience could contribute to solving real-world problems neurologically rather than behaviourally or cognitively unless fairly direct neurological intervention was the contribution.

It does seem, however, that that neuroscience research can *inspire* us to attempt what may very broadly be considered behavioural interventions that we would not have been inspired to try without the neurological data. The neurological finding that neurones could regenerate after stroke[6]. inspired behaviourists to attempt to rehabilitate stroke victims. *Even though* rehabilitating stroke victims is paradigmatically a behaviourist intervention and attempts are assessable from within the behaviourist paradigm it did, as a

---

[4]I suppose there is a problem in that administering some substance or other intervention on the brain might be considered - very broadly speaking - an alteration in stimuli. It does seem that there are many different ways that individuals can be disordered, however, so neurological data is well suited to finding out about what is going on for an individual even if it is less well suited for the population level economic case.

[5]And of course this applies to cognitive and behavioural interventions, too, though perhaps less clearly in marketing and economics.

[6]Or maybe it was that cognitive / behavioural function could be reacquired after stroke - I forget.

matter of fact, take *neurological findings* to result in behaviourists attempting and assessing behavioural interventions. If understanding how neurologically normal individuals work assists us in developing models about what is going wrong in the case of neurological abnormality; and if neurological models of what is going wrong in the case of neurological abnormality either directly or indirectly result in benefits to sufferers; then cognitive neuroscience has a claim to being of more or less direct benefit to sufferers[7]. Whether those interventions are more or less neurological or behavioural history seems to have shown us that it can take neurological finding to inspire new treatments whether those treatments are more appropriately regarded as neurological or behavioural.

---

[7]More than philosophy, obviously.

Chapter 4

Re-entrant processing

In the case of such longer-latency re-entrant attention effects in low-level sensory areas, do you think that this should be considered early selection, late selection, or something else? What would we need to know about the function of this late modulation to answer this question? What kinds of experiments could be done in the future to better understand this issue?

One dimension on which we can distinguish 'early selection' and 'late selection' is in terms of the amount of time (latency) between the onset of the stimulus and and the effects of attention. I shall call this the (T) dimension (for 'time'). There could be different notions of (T)early as follows:

$(T)early70$ which refers to latency of less than 70ms, $(T)early100$ which refers to latency of less than 100ms, (T)early...t which refers to latency of less than ...tms. We could go through a similar process to define different notions of (T)late. While theorists can agree as to whether one process is (T)earlier than another process it is important to note that theorists could disagree as to whether a finding indicates 'early' or 'late' effects of attention depending on what latencies they accept as 'early' or 'late'. If one theorist defines (T)late as a process that occurs after 100ms while another theorist defines (T)late as a process that occurs after 70ms then they are going to (seem to) disagree as to whether an effect at 80ms is late or not, for instance.

Another dimension on which we can distinguish 'early selection' and 'late selection' is in terms of how peripheral or central an anatomical structure is. I shall call this the (A) dimension (for 'anatomical'). There could be different notions of (A)early as follows: $(A)early thalmus$, $(A)early V1$, $(A)early V2$, $(A)early...$. We could go through a similar process to define different notions of (A)late. Just as with different (T) notions, different (A) notions could result in agreement that there are effects of attention at location a and agreement over whether one process is (A)earlier than another— but there could still be disagreement over whether this provides evidence for 'early' or 'late' selection depending on the particular notion of (A)early that one adopts. Finding evidence that attention affects V2 will only count as evidence for (A)early effects of attention if a theorist regards V2 to be A(early), for example.

Re-entrant processes are interesting because they show that the (T) dimension and the (A) dimension (as defined above) can come apart. We can agree that 300ms after onset of stimulus is (T)later than 100ms after onset of stimulus. Re-entrant processes thus are unambiguously (T)later than other effects of attention. Whether re-entrant processes are (T)early (simplicitor) will depend on which notion of (T)early one adopts, however (as illustrated in the first paragraph). We can similarly agree that V2 processing is (A)earlier than association area processing. Re-entrant processes thus are unambiguously (A)earlier than other effects of attention. Whether re-entrant processes are (A)early (simplicitor) will depend on which notion of (A)early one adopts, however (as illustrated in the second paragraph). At this point the issue seems to be whether 'earlier' should track the reference of the (T) or the (A) notion when they come apart (as they are found to do in the case of re-entrant processing).

At this point we may well start to wonder what the fuss is all about[1]. The dispute is starting to sound purely verbal and like this whole issue could be avoided if people simply described their finding (as in the third paragraph)

---

[1]Well, philosophers might be tempted to at any rate.

rather than worrying about whether the finding should be called 'early' or 'late'. In order to see the significance of the problem it will be useful to consider firstly how cognitive neuroscientists got to be interested in the issue of 'early' vs 'late' effects of attention, secondly how the finding problematizes feed-forward assumptions, and thirdly how the finding raises issues that have yet to be resolved.

The dispute seems to have originated in cognitive psychological theories of 'early selection' and 'late selection'. 'Early selection' theorists maintained that attention could affect early processing by which they seemed to mean the processing of stimulus features such as whether a speaker is male vs female, young vs old, speaking English vs Chinese, etc. In contrast, 'late selection' theorists maintained that attention affected processing of 'higher level' stimulus features such as processing for semantic meaning and / or encoding and / or retrieving from memory. The assumption seems to be that basic sensory features must be processed before more complex sensory features that then feed into motor systems (so a (T) / (F) feed-forward assumption, in effect). This processing is (plausibly enough) expected to take time and hence the interest in the temporal dimension as a relevant measure of the sensory and non-sensory functional distinction).

There is a schematic diagram in the textbook of how cognitive psychologists (roughly) envisaged this distinction mapping onto the brain. Subcortical areas were regarded to be 'early sensory' and temporally early whereas cortical areas (including V1-V4) were considered 'late non-sensory' and temporally late. This schematic differs from the way in which cognitive neuroscientists have come to distinguish (A)early from (A)late processing regions, however. In particular, the cortical regions V1-V4 are considered '(A)early sensory' processing areas by cognitive neuroscientists. The rationale for considering VI-V4 to be 'early' processing areas seems to be that those anatomical areas function to process the sorts of stimulus features that cognitive psychologists considered 'early', however. Our interest in anatomical dimension of 'early' and 'late' seems to be motivated by the assumption that information processing must follow anatomical pathways in the brain and that sensory

processing areas feed into association processing areas which feed into motor production areas (so a (F) / (A) feed-forward assumption, in effect).

It thus seems that another dimension on which we can distinguish 'early selection' and 'late selection' is in terms of how functionally low vs. high level the features are that are being processed. I shall call this the (F) dimension (for 'functional'). There could thus be different notions of (F)early as follows: $(F)early voices - vs - noises$, $(F)early male - vs - female$, $(F)early english - vs - chinese$, $(F)early....$ We could go through a similar process to define different notions of (F) late. Just as with different (T) and different (A) notions, different (F) notions could result in agreement that there are differences in the information that is being processed and agreement over whether one process is (F)earlier than another— but there could still be disagreement over whether this provides evidence for 'early' or 'late' selection depending on the particular notion of (F)early that one adopts. Finding evidence that attention affects the processing of sensory feature x will only count as evidence for (F)early effects of attention if a theorist regards sensory feature x to be F(early), for example.

We have considered already how re-entrant processing shows us that the temporal and anatomical dimensions are distinct. If we accept that the functional dimension is more important than either of these (as I think we should if we are interested in the prospects for an integrated cognitive neuroscience) then we need to know the function of the re-entrant processing. In particular, it would be helpful to know whether re-entrant processing consists in a fairly much unaltered re-presentation of the same informational content, or whether the content has been or is altered on the second run through. If we couldn't find a difference then we would seem to have some grounds for concluding that (F)early processing can occur at longer latencies[2]. If differences were found, however, then we may have some grounds for concluding that (F)late processing can occur in (A)early locations. The reason to think these are (F)late processes would be that they are incorporating information from association areas such that they may be better thought to

---

[2]Though inferring no difference on the basis of not being able to find one is problematic.

be functioning as association areas when they are activated at longer latencies. Since we have strong evidence for those anatomical areas processing (F)early sensory features if differences were found then one might be best to conclude that there is some evidence that a single anatomical location can be involved in both (F)early and (F)late processing[3]. I'm also interested in how the areas performing re-entrant processing deal with first entry processing of basic stimulus features when required to do so at the same latency as the re-entrant processes. I'm not sure if one is able to present two different stimuli 1 and 2 and see evidence of which is being neurally processed (or alternatively whether a finding would suggest that both are being processed concurrently on the same neural network). It might be that one kind of process takes priority[4].

---

[3]I suppose that if differences were found then we would still need to figure out what the re-entrant processes were doing. Another way of attempting to figure this out would be to attempt to disrupt or prevent the re-entrant processing and measure what subjects seemed unable to do in virtue of that disruption. This would seem to risk confounds with effects of the interruption compared with what it was that was being interrupted, however. Perhaps there are methodologies for sorting these issues out, unfortunately I don't know enough about them to feel like I have much that is sensible to offer.

[4]I would expect that sensory processing might take priority as people found that attentional blink stimuli were processed for their sensory features. I am unclear as to whether attentional blink is correlated with a stimuli not being re-entrant processed. It might be that re-entrant processing has something to do with iconic memory, for example.

Chapter 5

Representational contents

When data seems to conflict with a theory there are two things we can do. The first is to revise the theory so that it is consistent with the data. There are complex issues around how much alteration a theory can take and still count as numerically the same theory[1]. The second is to defend a theory from data by problematizing auxiliary assumptions since data doesn't support or dis-confirmed theory directly, but only by way of auxiliary assumption. In what follows I'll suggest a little of both. I'll suggest that both Cowan and Baddeley update the different kinds of functional / task distinctness that they posit in order to deal with behavioural dissociations and interference effects. Whether this will alter their identity as theories relies on the plausibility of my having interpreted the most significant difference between Cowan and Baddeley as that of whether the representational contents that are involved in the different kinds of activity are the same (Cowan) or different (Baddeley). I'll maintain that it is possible to defend Cowan's claim that working memory is mediated by long-term memory representations by maintaining that the neurological dissociations that have been found indicate task / activity independence rather than a difference in representational content.

---

[1]The issue of whether we have the same theory or not isn't as interesting as developing a good one. People do seem to get hung up over who was right and who was wrong, however.

This will basically involve analysing the localization assumption into two variants: Localization of representational contents (sensory cortex activity and motor cortex activity) and localization of function / task (more frontal activity). While it is possible to defend Cowan's claim about sameness of representational content in this way I do think that a more plausible view will consist in overlapping contents (same in some respects, different in some other respects) *all the way up*[2].

In cognitive psychology behavioural double dissociation is typically taken to be just the kind of finding that supports the claim that there are two distinct kinds of activity / task / function. Cowan's model posits different kinds of activity. There is a problem, however, in that there are different ways in which we can interpret the different kinds of activity that are posited by Cowan's model. One is that there are three kinds of *unstructured* activity as follows: declarative long-term remembering, actively remembering, attending remembering. If this is correct then finding double dissociations between each of these activities would seem to support Cowan's model rather than undermining it (though Cowan would need to posit further kinds of activity in order to handle other dissociation findings - as would Baddeley).

Behavioural double dissociations have been thought to be problematic for Cowan's model, however. The thought seems to be as follows: One way of understanding the claim that 'working memory is mediated by declarative long-term memory' is to understand it as a claim about the *structure* of tasks. So the thought is that rather than interpreting Cowan's model as making claims about *independent* tasks (where double dissociations are

---

[2]My view is motivated mainly by philosophical considerations, admittedly (and my view would probably be denied by most philosophers). Still, executive processing has enough problems to deal with without being gifted the central controller - most unfair! Enriched structured contents (explained a little later) allow us to avoid the central controller and also, I maintain, have better prospects for the development of a science of consciousness, though I don't have the space to get into that here. Can HM learn to run a maze (by the way)? If so Lashley's failure to find localization doesn't have obvious consequences for declarative long-term representations

precisely what should motivate the claim of task independence), Cowan is making a claim about *dependency relations* between tasks as follows: declarative long-term remembering, activation of declarative long-term remembering, attending activation of declarative long- term remembering[3]. On this way of specifying the kinds of activity it would seem that finding a person who was unable to perform declarative long-term memory tasks who was able to perform working memory tasks would be inconsistent with the model[4].

Another way of understanding the claim that 'working memory is mediated by declarative long-term memory' is to understand it as a claim about *representational contents* rather than *structure of tasks*, however. I think that the main point of difference between Cowan and Baddeley is that while both posit that there are different kinds of activity Cowan maintains that the representational contents that are involved in the different kinds of activity are the same whereas Baddely maintains that the representational contents that are involved in the different kinds of activity are different. On this latter way of understanding the most important difference between Cowan and Baddeley's models both of them seem comparably able to be modified to account for behavioural data (while retaining their identity as distinct theories). They simply need to posit different kinds of (unstructured) activity to handle the double dissociations that are found and task independence or overlap can explain dissociation or interference respectively.

---

[3]Though this would only be the case if we analyse 'x is mediated by y' into 'y is necessary for x' and thus the finding that 'not y and x' would dis-confirm the hypothesis. 'Mediated by' seems to be precisely the kind of fudge word that renders a claim immune from this kind of dis-confirmation, however. As such, even if we did read Cowan as committing to structured activities it is unclear whether finding dissociation between activities would dis-confirm the model.

[4]Though this would only be the case if we analyze 'x is mediated by y' into 'y is necessary for x' and thus the finding that 'not y and x' would disconfirm the hypothesis. 'Mediated by' seems to be precisely the kind of fudge word that renders a claim immune from this kind of disconfirmation, however. As such, even if we did read Cowan as committing to structured activities it is unclear whether finding dissociation between activities would disconfirm the model.

So, if we now turn to neurological data then it seems that we need to decide whether sameness or difference of local activation is going to be interpreted as indicating sameness or difference of task / function / activity or sameness or difference of representational content. The localization assumption that seems to be employed in cognitive neuroscience is that activity in the sensory cortices indicates representational content whereas activity in other regions (particularly frontal regions) indicates function or activity. It seems to be what is employed when theorists claim that such functions as 'attention' or 'consciousness' are localized in frontal areas whereas the representational contents of those functions are localized in sensory cortical regions. The urge to localize representational content in sensory cortices comes from finding neural tuning curves in those regions. There are significant problems in individuating content, however. Does a heightened amplitude in attended sensory processing indicate sameness of content, overlap in content, or difference in content? How do we decide? As we move into more frontal regions the urge to equate different activation with different content and sameness or similarity of activation with sameness or similarity of content seems to disappear. Now what we seem to want to say is that sameness or difference indicates sameness or difference of task, activity, or process. Cowan's model could thus be defended by interpreting the differences in neurological activation that have been found to reflect differences in functional activity rather than reflecting different kinds of content[5].

The desire to take activation in frontal regions to indicate activity seems to be partly due to our (currently) not having much in the way of information about representational contents in those regions, however. Philosophers probably have a lot to answer for when they say (falsely IMHO) that 'content' refers to the representation of stimulus features where those features are *outside the head*. We basically need an *enriched* view of content. While it will be easier to find that neurons are tuned to features outside the head rather

---

[5]We could also defend Cowan's model by maintaining that content can be distributed such that differences in neural activity don't reflect differences in content though this might be thought to be more radical.

24

than being tuned to the contents of lower level neurons (so re-representing) our failure doesn't imply that this isn't the case. I think that taking activation in those areas to indicate differences in activity rather than content is problematic[6]. A controversial issue (in the philosophy of language and the philosophy of mind) is whether we are best to carve activity such that the same content can feature in different activities where the activities might even have a complex structure (as Cowan seems to have been interpreted as doing): representing perceptually(p) or representing, consciously, perceptually(p) and so on. Or alternatively, whether activities are something that is better to be understood as *certain kinds of representational content* as follows: representing(perceptually, p) or representing(consciously, perceptually, p) and so on[7]. The latter view seems less suggestive of a central controller, however. While it does seem that at this point we would have given up on both Cowan and Baddeley's model it would seem fairly intuitive that contents can (sometimes) be the same (in some respects) and that contents can (sometimes) be different (in some respects) rather than taking extreme views as Cowan and Baddeley do (or interpreting the localization assumption differently for different neurological regions). It might seem counter-intuitive in reducing function to content, however (not sure how plausible or crazy that is going to seem[8]. It is also unclear to me how this model could be tested, though to be fair a full defence of Cowan's model would seem to result in a

---

[6]The inevitable question that seems to arise from taking frontal processing to indicate task or activity is 'what gets to allocate resources to task or activity, or what decides which takes priority?' Onset of stimuli is useful for time locking trials. The representation of task instructions might be useful for the content we view as 'central controller'.

[7]Of course the terms in the parentheses will need to be broken down properly into the contents. Individuating / carving contents up correctly is going to be a tricky matter. They will also need to be structured (to allow for inference / transformations etc).

[8]This comes up in the consciousness literature. Most grant that phenomenally experienced properties (e.g., subjective experience of brightness) entail certain kinds of representational contents but some maintain that representational contents won't entail phenomenally experienced properties unless you build phenomenality into the representational content. Building it in is a way of getting a two-way entailment (which might be thought to be necessary for an identity). I'm wondering how much a similar strategy could generalize back to other forms of what are typically taken to be activity such as working memory or declarative long-term memory or attention.

similar problem.

Chapter 6

Emotion

Affect program responses have been useful for bridging ethology and an-thropology[1]. Affect program responses are interesting because they share a number of properties in common that are useful for induction: They are fast, short lived, have distinctive patterns of (fairly) involuntary bodily and behavioural response, they are pan-cultural, and they have homologues in other mammals. There seem to be good prospects for explaining how af-fect program responses can be modified so as to be elicited in response to novel social and non-social cues. There also seem to be good prospects for explaining how affect program responses are modified by a variety of social and non-social reinforcements. One can thus investigate how non-social and social factors influence their production and expression both in humans and in other mammals. This seems a promising approach to grounding the va-riety of more culturally specific emotional responses in the biological sciences.

With respect to production humans seem hard wired (or at least prepared) to respond to non-social cues such as insects and heights with fear. We also seem hard wired (or at least prepared) to respond to social cues such as

---

[1]I know there has been some controversy over precisely how many there are and also some questioning of how significantly cross-cultural they are. The findings for Ekman's 7 do seem fairly robust, though.

expression of fear in another with fear. Both the non-social and the social cues that trigger an emotion can be modified (e.g., generalized to apply to different stimuli or overcome) as a function of learning. This can arise from non-social forms of feedback (e.g., nausea after eating a certain food), or from social forms of feedback (e.g., another responding to us with disgust when we eat a certain food[2].

While affect program responses are fairly automatic and involuntary, culture seems to influence not only whether an emotional response is produced in response to a stimulus, but also the intensity, and type of response. The finding that subjects from eastern cultures produce a display that is less intense than that produced by subjects from western cultures has often been interpreted as indicating that subjects from eastern cultures engage in 'suppression' of their emotional response. One might conversely interpret the finding as indicating that subjects from western cultures engage in 'magnification' of their emotional response, however. Whether the differences arise from different intensity of emotion or different display rules remains to be seen[3]. Cultural differences can also result in different emotional responses to the same stimulus as when personal achievement tends to elicit pride in one cultural group and humility in another. Similarly, which affect program response is elicited can depend on whether the behavioural stimulus was produced by a subordinate or dominant. There are also differences in the display rules in different cultural groups or sub-cultural groups (e.g., displays of violence are obligatory in some subcultures).

---

[2]Much to the consternation of the New Zealand Dairy Board who is trying to figure out a way of marketing dairy products to the segment of the Chinese population that don't have lactose intolerance but find dairy disgusting nevertheless.

[3]I think one should be wary of describing eastern subjects as 'suppressing' their emotion as the implication seems to be that this is the same variety of 'suppression' that has been found to be a maladaptive coping strategy. I'd be interested to know if intensity of SCR between the eastern and western subjects was correlated with the degree of display. If so then this would suggest that the cultural differences resulted in different intensities of emotions being produced rather than that the cultural differences were solely a matter of display rule.

Despite their utility for ethology and anthropology, affect program responses haven't turned out to be very useful (or recoverable) from the perspective of neuroscience thus far. This is surprising to me as there clearly are distinctive patterns in behavioural and bodily response for the different affect programs[4]. Unless we are simply missing something I suppose it must follow that the different affect programs are implemented on the same (or very similar) neural circuitry. We may have been suspicious about their being implemented on different circuitry at any rate since double dissociations between distinct affect program responses didn't seem forthcoming[5]. Perhaps they simply aren't as distinct as we had supposed on the basis of the behavioural evidence. It might be that instead of starting the story with affect program responses and building up to the socially constructed emotions and socially sustained pretences (where an emotional display is elicited because of its accepted meaning in a culture e.g., possession by a wild pig syndrome) we need to start the story further back with how physiological arousal and approach / avoidance build up to the affect program responses (preferably by way of valence). Someone looked at the difference between fear and anger production in rats (either Damasio or Le Doux). It seemed that when escape was impossible or the threat could be reduced by anger display then anger was produced, whereas when escape was possible or the threat couldn't be reduced by anger display then flight was produced. I can't remember how closely this was tied to neurological circuitry, however, or whether the only discovered dimensions of difference were the different features I listed. I do think that it is fairly important to recover something along the lines of affect program responses, however, even if they are implemented on the same neural circuitry, as they have turned out to be immensely useful for integrating the biological and social sciences of emotion. I'm not at all sure how SCR and fMRI correlations (that seem fairly non-specific) are going to be able to

---

[4]Perhaps closer to the motor cortex?

[5]Though perhaps this is still surprising as it has been found that you can get double dissociations out of a neural network that has been trained to compute two different tasks / contents. I'm not sure how many hundreds or millions or billions of trials that required, however.

play a similar role for neuroscience and social psychology[6].

The dispute over whether cognition and social cognition involve different contents / stimuli or consist in different processing mechanisms reminds me a lot of the dispute over whether endogenous and exogenous attention involve different contents / stimuli or consist in different processing mechanisms. Cues differ from symbols on a number of different dimensions such as where the stimuli was located (centrally vs peripherally) etc. It would be interesting to see what would happen if the dimensions of difference were systematically varied (morphing a cue into a symbol, in effect) in order to measure whether the effect on both neurological processing of the stimuli, and the pattern of behavioural response. The difference in stimuli is a matter of degree and it would be interesting to see whether neurological processing and the behavioural profile were similarly a matter of degree or whether they turned out to be more categorical. With respect to the distinction between social and non-social cognition it would similarly seem that we can morph uncontroversially non-social cues into cues that are uncontroversially social. I am supposing that the interest in the relationship between emotion and social cognition is thought to be a distinct problem from the problem of the relationship between emotion and cognition more generally. Whether this is so remains to be seen.

Returning to SCR one theory (controversial to be sure) is that the Capgras delusion is a reverse dissociation of prosopagnosia. Subjects with the Capgras delusion maintain that someone close to them (e.g., partner, child, canary) has been replaced by an impostor that looks 'just like' the original. It has been found that subjects who develop the Capgras delusion in response to

---

[6]I would think it would be preferable to keep ones distance from people who are proclaiming to have found the 'neural basis of racism' after such fiascos as the bell curve, the 'fundamental attribution error', the 'finding' that indigenous people couldn't perform modus ponens, and the last sociobiological effort that led to eugenics. Anthropology provides a way of keeping our generalizations from a very culturally specific subset of humanity in check.

cerebral trauma have a loss of SCR to familiar faces (and canaries, one supposes). I don't know that anyone has tested psychotic individuals with the Capgras delusion to see whether they similarly have reduced SCR. There has been some puzzlement over how the Capgras delusion differs from the Cotard delusion where people maintain they are dead. People with the Cotard delusion similarly have loss of SCR to familiar faces, though a lot else besides since it typically occurs in the context of untreated depression.

Marsha Linehan (University of Washington) developed a treatment for borderline personality disorder which is now considered paradigmatic of emotion dysregulation. Dialectical Behaviour Therapy (DBT) was found to outperform CBT and psychodynamic therapy[7]. It consists three components: mindfulness meditation, emotion regulation, interpersonal skills. She has said that future research will need to determine which aspects (or which components of which aspects) are responsible for its efficacy. One thing that is interesting about abnormal SCR in delusion is that it may turn out that that delusions are more appropriately regarded as disorders of emotion rather than disorders of cognition. DBT doesn't focus on reappraisal / cognitive restructuring so much as acceptance of emotional response. Since cognitive restructuring results in high drop out rates and low rapport with psychotic subjects (as it did for subjects with borderline personality disorder) applying DBT techniques to the treatment of psychotic individuals might be promising[8].

---

[7]Trauma re-experiencing is not a part of DBT. The thought is that these people struggle enough with whatever happened to them that morning. Asking them what happened that morning probably wouldn't count as asking them to remember their trauma. One would be expected to get an intense response nevertheless.

[8]Anti-psychotics are also often given to people with borderline personality disorder (or anxiety) for the purposes of muting emotional responses. I suppose it is a significant problem assessing the emotional responses of psychotic individuals compared to the emotional responses of psychotic individuals on emotion altering anti-psychotic medication. It is a strange situation indeed when running an efficacy study of therapy (and no medication) compared to medication (and no therapy) is considered unethical for a medication that was only approved because its side effects were 'less severe' than those discovered (thus far) to obtain to those of the previous generation...

Chapter 7

Language

Is human language unique among the social communication abilities of other species?

In order to assess whether human language is unique among the social communication abilities of other species we need to begin by characterizing the essential features of human language (or, more plausibly, a few central features). Then we can assess how many of those features are met, to which extent, by which other species. If we are interested in the uniqueness of human language in particular, then we would also need to characterize the essential features of other species communications (or, more plausibly, a few central features of each other species communications). Then we could assess how many of those features are met, to which extent, by which other species. We would then be in the position to see whether human language is an outlier and if so to what extent. Of course we haven't got anything approximating this at present.

In attempting a taxonomy of species on the basis of morphological features there is a problem in picking out what morphological features are relevant for the taxonomy. Whether two species are 'closer' than two other species

is highly sensitive to the morphological features that are chosen for the purposes of the analysis[1]. Evolution by natural selection was thought to help distinguish the relevant from irrelevant features for the development of a classification scheme[2]. In linguistics Chomsky's poverty of the stimulus arguments persuaded many that there was a gap between the learning history of human beings and the linguistic competence that they attained. There also seemed to be an obvious gap in the sense that even if we raise a member of another species such that it has a comparable linguistic learning history as a human the animal will fail to achieve the linguistic competency achieved by almost all human beings[3]. The obvious solution was to say that what was in the gap was innate and uniquely human. Chomsky's thought was that that was the content of the innate language acquisition device (LAD).

The LAD hypothesis provided a useful framework assumption / paradigm for linguists who could get on with the business of generating and testing more specific hypotheses as to the contents of the LAD. While this served to organize linguistics for a time eventually the field became dissatisfied with their progress and started to grumble that the LAD hypothesis was immune to support or dis-confirmation and they moved on to other things. Despite this, there are two features of human linguistic capacity that have inspired much interest with respect to their seeming centrality for what carves human linguistic and cognitive capacity off from the linguistic and cognitive capacity of other species. Those features are systematicity and productivity[4].

---

[1]A similar problem comes up when there are basically three different measures of bio-diversity such that it is jolly hard for us to figure out what it is that we are supposed to be maximising. Basically... Whatever measure gives us biggish mammals or things that are, were, or might be cute. Which measure we want seems to vary depending on which outcome we are interested in. I have a similar suspicion regarding which features of language / communication we take to be relevant.

[2]Though even here lineage trees vary considerably depending on assumptions regarding the rate of change and which kind of data you prefer when they conflict etc.

[3]Insofar as it is possible to raise a member of another species comparably to a member of our own which I suspect it is not given other functional differences and differences in normal infant / caretaker relations between species.

[4]Other features have been cited besides, but I'll focus on these. I think it is fair to

Systematicity refers to the feature whereby we can combine and recombine meaningful units in different systematic ways (so 'the cat is on the mat' and 'the mat is on the cat' share meaningful units). Productivity refers to the feature whereby even when we hold the vocabulary / meaningful units fixed, we can still generate an indefinite number of meaningfully different sentences / thoughts and also understand novel sentences / thoughts of others (e.g., 'I have 1 kumquat', 'I have 4,837,432,494 kumquats' etc). If we take this to be the best (or one of the best) candidates for what is distinctive about human language then the question becomes to what degree (if any) these features are present in the communications of other species[5].

One thing that we should be concerned about is that we seem to need a good account of the semantics that are employed by other species so that we can figure out the rules of arranging those (a syntax). I'm not sure that we do have anything like a complete account of the semantics that are employed by other species, however. I'm not even sure that we have anything like a complete account of the semantics of our own. While we understand some features of honeybee dances, vervet calls, and birdsong, it is unclear to me whether we know all, most, some, or next to nothing of the semantics they employ. This issue is complicated by other species using forms of communication that are difficult for us to assess, and sometimes even difficult for us to think to assess (e.g., utilizing frequencies that are hard for us to detect or using chemical trails or echolocation(?) etc). It wouldn't even be so very surprising if they were making semantic discriminations that we aren't picking up on because they are different from us with respect to what is salient

---

say that these are the best candidates for categorical uniqueness. I'm also shifting slightly from language to thought. I want to say that the reason we care about the linguistic abilities of other species is because we think it indicates something about their cognitive capacities. Trouble is that some others want to say that the reason we care about their cognitive capacities is because they think it indicates (sometimes by way of consciousness sometimes not) something about the appropriate moral attitude we should take towards them. I guess it depends on what you are interested in.

[5]I do worry about whether they count as three different features or two different features or one different feature. I suppose that will make a difference for just how unique human language is going to turn out to be.

to their interests in the niche that they find themselves in. If this is the case then we wouldn't have much of an idea what (if any) syntactic rules govern their semantics. One way around this issue is to see if we can find evidence of their being trained to distinguish recursive from non-recursive syntax in tasks that we set. Since syntax is meant to be (fairly) independent of content it might be thought that in order to have syntactic recursion a species would need to be able to identify recursive syntactic patterns even if the semantics is meaningless to them.

I am a little concerned about Hauser's study insofar as the stimulus consisted in human speech sounds rather than monkey vocalizations, however. It seems that the song birds were at an advantage in having a stimulus that was a song bird vocalization. While syntax is often thought to be completely independent of content I'm not terribly sure about this. Logic is also thought to be independent of semantic content even though the paradigmatically syntactic boolean connectives are given semantic truth tables[6] People tend to do better at detecting valid inferences when they have the form with the content removed. Conversely, people have been found to perform better on socially relevant tasks than they have been found to perform on what is effectively the same task stated abstractly. It might similarly be the case that recursion is easier to identify when the semantics that the recursion is operating over are meaningful (or familiar) to the subject. It might also be the case that species perform best when we set them tasks that make sense to them given their ecological niche. It is a little like how some indigenous people wouldn't perform modus ponens because they refused to speculate about something they had not seen. This didn't indicate that they *couldn't* speculate or think hypothetically, however. Rather it indicates that we need to be careful in how we ask them to think hypothetically. A similar thing arose with dominance hierarchies affecting what tasks monkeys would perform in the theory of mind literature. Similarly, it it unclear to me what the experiments show

---

[6]Mathematical functions might be a better candidate for being a content-less purely formal / structural property, but I'm not sure about this.

us (if anything) about syntactic abilities. I'm not sure how pattern recognition of sounds counts as syntax (maybe it is just that songbirds are more musical). It might well be that I simply don't understand what syntax is supposed to be, however...

Chapter 8

Plasticity

The somatosensory system is organized in a hierarchy of populations of neurons[1]. The neurons in the subcortical and primary somatosensory cortex (S1) are topographically arranged at each point in the hierarchy such that neurons that are adjacent (more or less) selectively respond to stimuli in adjacent regions of the body. It has been found that the somatosensory cortex is plastic, in the sense that regions that used to respond selectively to mechanical stimulation of a particular digit can come to respond selectively to mechanical stimulation of an adjacent digit if that first digit is removed[2] It has also been found that mechanically stimulating a particular digit or digits repeatedly can result in cortical magnification of the region/s processing that information.

The motor system is similarly arranged in hierarchical populations of neurons. There is a topographic arrangement here, too, with adjacent neurons resulting in movement or muscular contraction of (more or less) adjacent regions of the body. The representational contents of the motor cortex (M1) seem to have been the subject of much more controversy than the represen

[1]As are the other sensory systems. I'll focus on the somatosensory system here, however.

[2]I will restrict myself to the issue of plasticity in the central nervous system of adult primates.

tational contents of the sensory cortices, however (at least that is the impression that I got from the textbook especially regarding visual processing)[3]. We need to have some idea of what the neurons in M1 represent order to assess whether the representational contents of those neurons have changed or whether they are likely to be able to change in a way that is comparable to the plasticity that has been found in S1.

Most of the controversy over the contents of M1 seems to be over whether the neurons represent *movement* or whether they represent *contraction of muscles*. There seems to be evidence for both where stimulating single neurons results in muscular contraction, whereas stimulating a region of adjacent neurons results in movements such as grasping or grimacing. One concern with regarding individual neurons to code for muscular contraction was that the tuning curves of individual neurons in M1 were fairly coarse grained with respect to their selectively (or not very selectively) being activated prior to a range of different movements. It was unclear to me whether the different movements involved an overlap of muscular contraction that corresponded to the overlap of individual neuronal activation, however. If this were found to be the case then the coarse grained tuning curve in relation to movement wouldn't be evidence against individual neurons having a more refined tuning curve in relation to muscular contraction.

In favour of M1 neurons coding for movement it was found that direction of movement could be predicted from averaging the activity of the neuronal

---

[3]I've been struggling with the a-symmetry between the sensory and motor systems where sensory systems representational contents are given by the contents of the systems that are temporally prior (typical causes), but where motor systems representational contents are thought to be given by the contents of systems that are temporally later (typical effects). We can predict the contents of the visual system fairly well with respect to retina, LGN, and V1. I'd be interested to know whether we are able to retrodict the contents of the visual system comparably well with respect to V1, LGN, retina. If so, then the a-symmetry probably doesn't matter. If not, then I'm simply not sure whether or not this is an important a-symmetry that will have further ramifications (e.g., for plasticity or for multiple realizability of representational contents) or not.

population and that direction of arm movement, in particular, could be predictably altered by temporarily anaesthetizing a subregion. The direction of movement wasn't able to be predicted from the activity of individual neurons, however, as the tuning curves are too coarse with respect to movement, as we have seen. I would be interested to know whether the monkeys were able to correct their direction of movement over a number of trials in the face of temporary anaesthetizing a subregion of M1. If so, then this would seem to be an interesting case of behavioural / movement plasticity as a function of (temporary) damage and training where the representational content of the anaesthetized region is presumably replicated (or compensated for) in some way by the surrounding region.

Perhaps the disagreement over the representational contents of M1 can be sorted if people are clear about whether they are talking about the representational contents of individual neurons in M1, or whether they are talking about the representational content of the population (or of some sub-population) of neurons in M1. With respect to what the neurons in M1 are really coding for, that might turn out to be about as interesting as whether an ambiguous figure is really a duck or really a rabbit (which is to say not at all)[4]. It might be that whether one is interested in how M1 is tuned to either muscular contraction or movement determines whether one is (or should be) more interested in individual neurons or the population of neurons. I really don't see why this isn't a similar issue / problem (or perhaps non-problem) for the contents of sensory systems (including the visual system).

The relationship between individual tuning curves and population tuning curves is going to be complex. The population is going to supervene on the individuals. That is to say that there cannot be a change in the (tuning curve) properties of a population without a change in the (tuning curve) properties

---

[4]Though it could be that I'm missing some of the significance of the data, here.

of the individuals that comprise the population)[5]. The population is also going to be multiply realizable. That is to say that there can be a change in the (tuning curve) properties of the individuals that comprise the population without a change in the (tuning curve) properties of the population[6]. Some cognitive neuroscientists (e.g., Bechtel) deny multiple realizability. Still, it does seem that motor programs may be multiply realized by movements and that movements may be multiply realized by muscular contractions. Multiple realizability together with supervenience (which seems fairly obviously to apply to compositional relations such as populations and the individuals that comprise them as well as to muscular contractions, movements, and motor programs) together seem to imply that there will be more plasticity at the ontology of lower levels than at the ontology of higher levels, however. Despite this, it seems fairly obvious that motor programs such as signing ones name or pressing a lever are very plastic on the basis of behavioural evidence that they can be acquired in adulthood. It would thus be highly surprising to me if movements weren't found to be plastic (my argument requires them to be even more so). Whether the plasticity is similar to that of the sensory system remains to be seen.

The issue then becomes one of what constraints there may be on plasticity. While motor programs may be massively multiply realizable with respect to muscular contractions, movement will be multiply realizable, though less so. Topography might well limit the tuning curves of individual neurons with respect to the connections they bear to inputs (in the case of sensory systems) and outputs (in the case of the motor system). If we grant that transduc-

---

[5]Gaining or losing members counts as a change in the individuals.

[6]Supervenience and multiple realizability are also thought to be features of the software / hardware distinction and are likely features of the relationship between muscular contraction, movement, and motor program responses. Both of these relations are a-symmetric, which gets me worrying about a-symmetry between the sensory and motor systems, again. On a slightly different note I'm used to a similar distinction between sensory, perceptual, and conceptual representational contents. I didn't appreciate that they were going to be lumped together until Cowan. If I had have appreciated it in the perception unit I would have had much to say about it...

ers and lower motor neurons are non-plastic in the sense that lower motor neurons don't attach themselves to different muscles even if the connections to fibres within a muscle may vary, then it might be that there are different degrees of plasticity at different levels of the processing (or individual / sub-population / population) hierarchy. Still, my argument bothers me because it seems to imply that individuals (or low level processing) will be more plastic than population (or higher level) processing. Something other than the motor system seems to be backwards (or maybe upside down). Still, I did read that individual neuron tuning curves vary over time, anyway. It may be that it is easier to fire with proximal rather than distal relations and that the plasticity of individual neurons in particular is constrained by what ones relatives are doing.

Chapter 9

Evolution and individual difference

What is the role of evolution in the expression of individual differences in cognition?[1]

It has been mathematically proven that if there is individual variation in traits or variants of traits, heritability such that offspring resemble their parents more than the parents of others, and differential fitness such that individual reproduce at different rates, then the proportion of the population with a trait or variant on a trait will alter over time. That is simply to say that if those three conditions obtain, then evolution by natural selection will occur. These three conditions are not only thought to be sufficient for evolution by natural selection, however, they are also thought to be necessary, or *defining of what evolution by natural selection is*. As such, individual difference is a precondition for evolution by natural selection[2].

---

[1]I'm going to argue largely from the principles of evolution by natural selection as I think the implications are often misunderstood - often with concerning consequences for social policy.

[2]It is worth distinguishing the claim that those three properties are instantiated in the world from the stronger (and probably false) claim that evolution by natural selection is the main mechanism for speciation. The intelligent designers are correct that there is

Genetic mutation is the mechanism that produces variation[3]. While most mutations are deleterious to the organism, occasionally mutations result in what Dennett refers to as a 'Very Good Trick', or adaptation[4]. The most plausible version of the thought is that a trait or variant on a trait might give such an advantage to the organisms that possess it that the trait or variant can be driven into fixation in the population. Language seems to be like that for *Homo Sapiens* in the sense that language is very robustly acquired by almost all across a diverse range of environments. It is important to note that adaptations are doubly relative, however. Firstly, what has highest fitness in one environment might have lowest fitness in another. Secondly, what trait or variant has highest fitness in a population is dependent on what other traits or variants are found in the population[5].

Some have conceptualized evolution by natural selection as a mechanism that can *produce* diversity rather than being a mechanism that produces homo-

---

much debate over the latter. It is unfortunate that these two distinct issues are often conflated.

[3]There is controversy over the units that natural selection operates on. Some candidates are: genes, morphological or behavioural characteristics, sub-populations or groups within a species, species, memes. It isn't quite clear what would play the role of genetic mutation for these other units. It is understood that there is a gap between genes and the expression of genes in morphological or behavioural characteristics, however.

[4]The 'Adaptationist Assumption' has come under a lot of fire recently, so it is possible that Darwin didn't go there for very good reason.

[5]Darwin was very clear in maintaining that it was senseless to attempt to rank different traits for their adaptive value *between species* . This is because different species occupy different ecological niches (e.g., is it better to have gills or lungs?). Despite this, we might think that we can rank different species with respect to how adapted they are to their own environmental niche. In particular, it would seem that species that are close to extinction aren't particularly adapted. Ants probably outperform *Homo Sapiens* with respect to diversity of environments that they occupy and prevalence around the globe. Now we have a decision: We can accept this or we can say that we aren't much interested in adaptation after all (because it doesn't seem to be working to justify our intuition that we are better somehow). Not a lot follows from being better adapted (unless extinction is imminent), however. In particular: Ants are not busily evolving to be more like us, and we are not busily evolving to be more like ants.

geneity (by way of fixation) by operating over diversity. The clearest example of this between species is thought to be the role of evolution by natural selection in speciation. Our best biological taxonomy of species distinguishes them according to lineage and thus interbreeding is more important for delineating species membership than morphological or behavioural characteristics. Evolution by natural selection is thought to operate on populations where populations are collections of interbreeding individuals. In order for speciation to occur a population needs to be subdivided into sub-populations where there isn't interbreeding between the sub-populations. The lack of interbreeding prevents heritability operating between the sub-populations and thus variations that arise in one population can't be driven into fixation across both sub-populations unless the variation independently arises in both groups. What variations arise in the sub-populations are likely to vary. What variations are adaptive in the sub-populations are thus also going to vary - both as a function of the different variants present in the sub-population and as a function of the different niche affordances of the two sub-populations - since the main mechanism for producing sub-populations is going to be geographical separation. A variety of extra-evolutionary mechanisms referred to as 'drift' (e.g., meteor strike, falling trees etc) are thought to be responsible for dividing populations into sub-populations. Once we have sub-populations then evolution by natural selection can (over time) result in the sub-populations being *unable* to interbreed for genetic or morphological rather than merely geographic reasons. While there is controversy, I think that in the case of speciation extra-evolutionary processes are responsible for the variation that is generated within a population and for the heritability of those variants (where geographical isolation prevents that). Speciation thus doesn't undermine the notion that evolution by natural selection operates over diversity to produce homogeneity[6].

---

[6]And thus extra-evolutionary processes are more significant for speciation than evolution by natural selection. The intention / mandate of an intelligent designer doesn't count as drift because the intention / mandate of an intelligent designer is not a scientific mechanism, however.

When a trait or a variant on a trait has been driven into fixation then we have a fairly clear case of adaptation or function for the trait or variant on the trait - in that species, in that environment. When variation remains and there isn't a clear shifting of the relative frequency of the trait or particular variants in subsequent populations then it would seem that we simply don't have a clear case of the trait or variant of the trait being more adaptive or functional than other traits or variants[7]. The persistence of difference in such things as IQ scores shows us that the differences aren't differences of evolutionary adaptation or function. Indeed, it is more surprising that people would think that they would be relevant as the evolutionary notion of adaptation or function has to do with the *relative number of offspring* and it is very clear that we aren't attempting to calibrate the findings of intelligence tests with relative number of offspring! Evolution by natural selection and the related notion of expected relative number of offspring thus doesn't seem to have anything to do with the diversity of intelligence found in intelligence tests. Finding diversity in intelligence with no clear shifting in the intelligence exhibited by subsequent generations shows us that individual differences in cognition aren't being subjected to evolution by natural selection since evolution by natural selection would result in greater homogeneity in scores rather than persisting diversity. Evolution by natural selection and the evolutionary notion of adaptation, function, and dysfunction are often misapplied. The misapplication seems most often to based on a misunderstanding of the principles of natural selection.

A similar situation has arisen where theorists have maintained that the difference between a disorder or disease and a mere problem in living is that evolutionary dysfunction is necessary for disorder or disease. The thought is

---

[7]There is a great deal of controversy over whether intelligence is increasing over time. How one wants to define intelligence seems to depend on where one is situated in life more than anything else. Is it more intelligent to be able to hunt fish in a kayak in freezing artic conditions or to perform *modus ponens* on arguments where the premises are false? Is it more intelligent to create bombs with the power to destroy the conditions ones species requires for life or to be able to survive such radical environmental change (e.g., cockroaches)?

that science will discover whether a certain condition involves an evolutionary dysfunction or not and that whether a condition is a disorder or a disease is thus something to be discovered by science. There are two problems: Firstly, evolutionary function and dysfunction are not the relevant notions for bio-medicine. While there is near consensus on the 'disease' or 'disorder' status of conditions that threaten survival increasing expected reproductive fitness simply is not the aim. Fertility treatment is one of the most contested areas and we often care about disorders occurring after reproductive age. Secondly, the relevant notion of dysfunction isn't *discovered* it is simply *assumed* . We begin with the phenomena that is of interest. We then describe the causal processes involved in its production. If we have the intuition that the phenomenon is a dysfunction then we describe the causal processes as dysfunctional. If we have the intuition that the phenomenon is not a dysfunction then we describe the causal processes as mere differences. Whether the causal mechanisms involved in the production of homosexuality were were regarded as dysfunctional or not depended on whether the morphology was regarded as dysfunctional or not, for example. I am very concerned about our failure to realize that our values drive our judgements of function and dysfunction both in the natural sciences (outside of evolutionary biology) and in social policy. It strikes me that there is a similar failure to see the values that are behind our efforts to rank such things as adaptation or cognitive capacity. While there is an objective notion of evolutionary function and dysfunction this notion really doesn't seem relevant for our project of ranking individual difference in cognitive capacities either between species or within species. While Homo Sapiens really are 'very peculiar primates' in a number of respects (see, for instance, Sterelny), the project of ranking difference seems to have more to do with imposition of value than scientific discovery[8].

---

[8]Fortunately this obsession with ranking individual difference in cognition seems largely restricted to individuals occupying a specific socio-economic niche in the USA and it doesn't really seem to have caught on elsewhere. While differences between species and within a species are interesting, I have immense difficulty with appeals to evolution by natural selection to justify rankings that are consistent with our values, especially when it seems to involve a misunderstanding of the principles of evolution by natural selection.