18.B: the ***Philosophyldeas*** overview of

Mechanics of Thought

All familiar thought occurs in brains. How the brain does this, employing neuron connections and brain chemistry, is the province of physical sciences, but the role of thought in philosophy is too important to wait for distant results which may never arrive. Psychologists also tackle the realities of thought by empirical methods, and philosophers should probably attend more to their findings. Philosophy, though, has its own contribution to make, usually by inferring the general structures needed to produce the patterns of thought that emerge at higher level.

Psychology was a largely introspective activity until the advent of behaviourism, which focused on what was objective and could be measured. But such strict empiricism missed out crucial interior facts about thought, and the commonest modern view is 'cognitivism', which sees minds as means to acquire knowledge and process information. A typical strategy of psychology is to analyse 'personality traits'; a standard proposal says these are agreeableness, extroversion, neuroticism, conscientiousness and openness to experience, and possibly honesty and humility. Most revealing for philosophy has been revelations about hidden motivations, such as in moral dilemmas. Research suggests that philosophers are too optimistic about knowledge gained from self-awareness, when people often contradict their apparent character, and reveal a darker side, often understood by researchers in evolutionary terms.

An ancient view of reality proposed that its ingredients fell into a few distinct **categories**, perceived by us, and that our thinking and language is thus structured by these categories. A standard list offered ten categories: substance, quality, quantity, relation, action, passion, place, time, state, ands activity. This list is guided by the world, but a later approach started with how the mind must be structured to experience the world. The mind is said to structure its judgements by four types of category (quantity, quality, relation, modality), each one having three modes. Such theories are highly speculative (though some have claimed to deduce them), and hard to test, but the possibility that thought is categorial in this way must be addressed, both by metaphysics, and by explanations of thought.

A modern approach says the brain has evolved in leaps, each adding a 'module' which enhances the range of available thought. This '**modularity of mind**' theory sees thought as teamwork, between separate processing systems. Each one has its own concepts, operating privately, with links to the rest of the brain. The language module is the most obvious, based on the innate requirements in any child learning a language. There may also be modules to deal with reading other minds, with biological understanding, with relating to physical events, and with perception. We each experience an inner 'personal assistant', giving unbidden reminders of tasks to perform and places to be. It is presumed that a controller (involving a rational self with a will) is needed for co-ordination and decisions. This theory mainly focuses on modules processing different types of information, but an alternative view sees the components of the mind as active '**drives**', struggling for dominance, and more fluid in their relationships.

Computers must relate both to human users and to silicon chips, and 'machine code' is employed by the hardware to do this. It is suggested that brains may have a similar hidden '**language of thought**' (dubbed 'mentalese'), which is roughly the same in all brains, whether they speak English or Mandarin. Unlike ordinary language, thought is never ambiguous, and a separate inner language can explain this fact. Theorists connect this notion of unambiguous thinking in an inner language with theories about 'propositions', and the concept of 'logical form'. If mental events such as beliefs and desires have a structure, then something like mentalese is needed to explain it.

There is little direct evidence for mentalese, but it offers a plausible way to link abstract thought to biology, especially if we say that **representations** of the world are constructed in this hidden language. Higher-level thought might then be a sort of computation, using these coded representations. A problem is that for mentalese to be a language it needs not only a grammar, but also a set of privately meaningful **concepts**. Since thinking depends on these concepts they need to be largely innate, which might result from natural selection for concepts like 'tree' or 'animal', but not for 'phone' or 'cake'. Rival accounts of the low-level way in which thought is implemented focus either on forms of **imagery**, or on types of internal **maps**.

Another suggestive analogy is the structure of computer databases. Thought may be built around '**mental files**', like folders in a computer system. The word 'elephant' makes you to open a mental file, and await further statements. You will respond to 'elephants are nice' with 'not always', by pulling nasty elephant incidents from the file. A mental file has a 'label' (usually a word), which connects to your acquaintance with the topic. There is a 'core' to the file, of basic facts, and this could be expressed by a definition. There is then 'peripheral' information, arising from links with other files. There may be a single type of file, with many variants, or several specific types of file, such as 'indexed' (for files in other minds), or 'encyclopaedic' (for collections of general information).

The mental file theory offers excellent explanations for some puzzling cases about language. A file might have two labels (Sam Clemens/Mark Twain); two files may be merged after a discovery (Morning Star/Evening Star); two files may share a label (Washington); one label may have two forms (London/Londres); and one label may lead to two files (Paderewski, as president, and as pianist). Most files are clear enough, but confusions inevitably arise in these tricky cases. If the mind is structured around files, that also clarifies our puzzles about universals. Obvious files are for single items, or distinct types, but some are much vaguer, and it is not clear whether the word 'thing' has a file.

Another line of enquiry is to see whether thought can be implemented on modern inorganic **machines**. The simple idea of a computer is a 'universal machine', reading symbols to move between states. These activities can then be interpreted as we wish. If thought is seen as functions, rather than physical events, then a machine performing universal functions can emulate thought. Deduction and representation are captured well in this system. It was predicted that eventually a machine would pass a test of being largely indistinguishable from a human in a blindfold conversation, and this has been achieved. However, a long list of problems make further progress daunting. Can a machine be conscious, have values, and cope with background assumptions, jokes, generalities, ambiguities, creativity, and relationships? Machines illuminate thought, but do not yet reproduce it.