Lecture 2: Attention

[Disclaimer: These informal lecture notes are not intended to be comprehensive - there are some additional ideas in the lectures and lecture slides, textbook, tutorial materials etc. As always, the lectures themselves are the best guide for what is and is not examinable content. However, I hope they are useful in picking out the core content in each lecture.]

Part 1: Definitions

The quote by William James at the start of the lecture is used to motivate three key questions. Here it is again in full:

> Attention is…the taking into possession of the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought.

The three questions posed in the lecture are:

- *How many* things are we attending to? ("one" or "several")
- *What kind* of things are we paying attention to ("objects" or "thoughts")
- *What controls* our attention? Do we direct it or does the world "take possession"?

List of different distinctions

From these three questions, we obtain several quite distinct aspects to attention. By considering the "how many" question, we arrive at the distinction between:

- **Focused attention** (a.k.a. selective attention)... when the goal is to attend to one thing and ignore all the other things
- **Divided attention** (a.k.a. multitasking)... when the goal is to keep track of many things

The "what kind" question leads us to consider the difference between

- **External attention**... when the *focus* of attention is something in the world.
- **Internal attention**... when the focus of attention is something you're thinking about

External attention is usually subdivided by sensory modality: **visual attention** is when you're paying attention to something you can see, **auditory attention** is directed at things you can hear, and **cross-modal attention** occurs when attention is directed to multiple senses at once (e.g., to the smell and taste of a cup of coffee)

Finally, the "what controls" question leads us to consider

- **Active attention** (endogenous control)... is when we exercise "top down" control to achieve our goals
- **Passive attention** (exogenous control)... is when something in the world imposes "bottom up" control, demanding our attention

Examples

This list of attentional "types" provides a pretty effective classification system for attentional phenomena. Consider the following examples:

> A car backfires in my street, causing me to startle and look out the window

The attention here is:

- *passive*, because it is being controlled by the world
- *external*, because it is directed at something in the world (specifically, something *auditory*)
- *focused*, because it is directed at a single thing

Part 2: Phenomena in auditory attention

A lot of the early work in attention focused on auditory attention, apparently because that was easier to study with the technology of the time. A lot of this work is inspired by the so-called **cocktail party problem** - trying to pay attention to a single "source" (e.g., one person speaking) when there are many different sources competing for your attention (e.g., it's a very noisy room and there are many people talking at once)

Much of this research relied on a paradigm called the **dichotic listening task**, in which a participant wears headphones and two different auditory signals are presented, one to the left ear and the other to the right ear. The task is to *shadow* the content in one ear (e.g., right ear), by repeating aloud what you hear in the target (right) ear, and ignore the content in the non-target ear (left).

[Empirical result] A typical pattern of results (e.g., Moray 1959) is that people are very good at the task, shadowing the target signal successfully. However, when tested, they remember very little of the content presented in the non-target ear. They can perhaps recall whether it was a male or female voice, and a few other "low level" perceptual features, but semantic information (i.e., the content!) seems not to be recalled.

[Theoretical idea] These kinds of findings led Broadbent (1954) to propose an **early selection** view of attention, sometimes referred to as "Broadbent's filter". The idea is that the perceptual system does some "simple" perceptual analysis of every signal, detecting things like pitch, loudness, frequency, etc. On the basis of this *preattentive analysis* the filter selects a single signal to pay attention to. Only this selected signal is processed further, and as a consequence we are only aware of the semantic information (i.e., meaningful content) within the attended stream. This idea explains the basic empirical results described earlier, but it has some problems.

Early selection theory predicts that we use low-level perceptual features to select the target for attention.

Yet there is evidence that people use semantic information to do so. It also predicts that we don't process the semantic content of the unattended stream, yet that also seems not to be correct.

[Empirical result] Here's the relevant findings from the lecture:

- Gray & Wedderburn (1960) - when the semantic content switches from one ear to the other, people follow the switch, preferring to shadow the semantic content. This makes a lot of intuitive sense, but it is inconsistent with early selection theory because we shouldn't be able to do this if the semantic processing only occurs *after* attention has been allocated.

- Lewis (1970) - when the unattended stream has contains information that is very semantically similar to the attended stream (e.g., two words with the same meaning but very different sounds), the unattended stream interferes with the shadowing task. This is known as **semantic interference**

- And of course, this was actually sort of well known from the beginning. Even Moray (1959) reported the fact that there are some special stimuli (e.g., your own name) that you will automatically (i.e., passively) attend to even if they appear in an unattended stream. To the extent that this is a *semantic* phenomenon (i.e., it occurs because your name is meaningful to you), this is an example of something where semantic content plays a role in the allocation of attention

[Theoretical interpretation] There are two alternative theories that were discussed that are able to account for these findings.

- **late selection** (Deutsch & Deutsch 1963) claims that all signals are processed almost entirely, including all the semantic information, it's just that we're often unaware of this. The claim is that because working memory capacity is limited, we're only able to "hold" a small amount of information in mind at once, so we're only "aware" of the attended stream.
- **attenuation theory** (Treisman 1964) proposes that sometimes we use early selection and sometimes we use late selection. Specifically, the claim is that we process every signal up until the point that we are able to determine whether it's relevant (i.e., is this the target signal I'm trying to attend to?). As soon as it becomes clear that this is not the target signal, processing stops

In the empirical literature there are follow up studies that seek to distinguish between these possibilities, but they're beyond the scope of this lecture.

Part 3: Phenomena in visual attention

In the third part of the lecture we asked whether the phenomena listed above are idiosyncratic to audition, or whether they are general phenomena that could be found in any sensory modality. We discussed two studies in particular.

- Rock & Guttman (1981) devised a visual analog of the shadowing task, in which people were shown two shapes drawn over the top of one another, one red and the other green, and the task is to rate the aesthetic appeal of the target "stream" (e.g., the red shapes) while ignoring the other stream (e.g.,

green). Participants were presented with many of these trials in quick succession. At the end, they were tested on their memory for whether they'd seen various shapes. Mirroring the typical results in auditory tasks, people had very good memory for shapes that appeared in the attended stream (red) and very poor memory for those in the unattended stream (green)

- Tipper (1985) pushed this analogy further, demonstrating a **negative priming** effect in visual attention that very closely mirrors the semantic interference effect in auditory attention (see above). The methodology was very similar to Rock & Guttman, but used illustrations of familiar, meaningful objects. The key result is that when the previous item in the unattended stream (e.g., a green wolf) is semantically related to the current item in the attended stream (e.g., a red cat), people are slower to respond to the current item. What this suggests is tha semantic information in the unattended (green) stream is in fact being processed. Whether we realise it or not, we are detecting that there is a wolf in the green stream and are "suppressing" our response to it because it belongs to the unattended stream. However, what this does is suppress our willingness to respond to wolves, cats, dogs, etc; so we are slightly slower to respond when the next item in the attended stream happens to be a red cat.

The key take home message from this is that there are some remarkable similarities in how attention operates across different sensory modalities.

Part 4: Visual search

The final part of the lecture discussed "visual search" tasks, in which the goal is to find a "target object" hidden among a collection of distractors (e.g., find my child among the crowd of children in the playground)

[empirical results] When the target is defined by a specific feature (e.g., colour) it seems to **"pop out"**. Attention is automatically (i.e., passively) drawn to the target item. The **set size** (i.e., number of distractor items) makes no difference to the search time. This phenomenon is quite general, and doesn't depend on what kind of feature differentiates the target item: colour, shape, size, orientation, motion, depth, all produce pop out effects.

In contrast, when the target does not possess any unique features, there is no pop-out effect. For instance, if we need to find a red horizontal rectangle in a field of red vertical rectangles and blue horizontal rectangles, you need to make use of *both* features (i.e., orientation and colour) to solve the search problem. Search is slower, and now the set size matters: the more distractors there are, the slower you are to find the target.

[theoretical interpretation] The explanation for this proposed by Treisman (1986) is referred to as "feature integration theory". The idea is that the perceptual system has many different "feature analyzers" that detect perceptual features (e.g., red, blue, horizontalness etc). These operate quickly and in parallel - so if you can solve a visual search problem using only a single feature, then set size is irrelevant because the *parallel* nature of the feature analyzers means that you're processing every part of the visual input at the same time.

However, the feature analyzers are distinct from one another. Just because one analyzer has detected

"redness" at a particular location and another has detected "horizontalness" at the same location doesn't mean that we automatically *bind* those two pieces of information together into a unified representation of the object. In order to do this **feature binding**, we need to direct *attention* to the location in question. Because attention is a slow, serial process (i.e., does one thing at a time), any visual search problem that requires feature binding (i.e., we need to use multiple features to solve it) will not produce a pop out effect, and visual search time will be slower when the set size is larger

[empirical data] One piece of evidence for feature integration theory (FIT) is the **illusory conjunction** phenomenon. FIT predicts that Ffature extraction occurs automatically and in parallel; but object recognition requires feature binding, a process that requires slow serial attentional processing of stimuli to be done accurately. If this is not allowed (e.g., stimuli are presented too quickly for attention to come into play), then errors in binding will occur and will be based on features extracted automatically during early perceptual processing. See lecture slides for illustration.