•2• *A framework for cognition*

Levels of mental representation

In the previous section we have seen that sensory, visual information about the world needs to be interpreted as an object based structure for us to make any sense of it. Both visual and object structures are mental representations, but at different levels of information. The visual level is derived from the raw sensory data obtained from the eyes, whereas the object level is derived from the combination of the visual representation and the perceiver's knowledge and experience of the world. This means that there must be a set of mental processes that convert the visual representation into an object representation, and that there must be other processes that allow memory to influence the object representation.

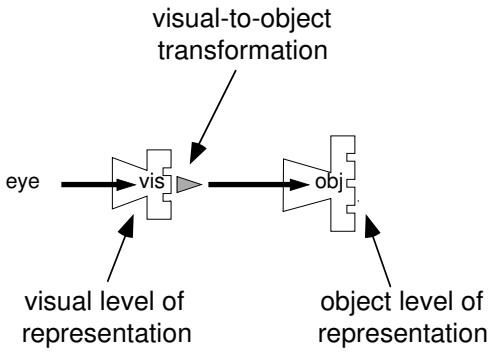

Fig 2.1: sensory information is transformed from a visual level of representation into an object level

The shapes in Figure 2.1 look odd, because they are parts of a larger model of human cognition, which we are going to describe piece by piece. When all of the parts are in place, their shapes will become more meaningful. The part of the overall model shown in this figure illustrates the process of perception that we have described so far. Sensory information about the world is detected by the eyes, and turned into a visual representation that contains a wealth of detail about colours, shades, contrasts, angles and edges. A mental process then 'interprets' this information, transforming it into an object representation, which contains information about lines, shapes, depth, position and orientation.

Sensory information from the eyes forms a **visual representation.**

Perceptual information is contained in an **object representation.**

It is important to remember this distinction between the sensory level of information in the visual representation, and the perceptual level of information in the object representation. One advantage of making this distinction is that it helps us analyse what people will *subjectively* think about a display design (their object representation) as well as what is *objectively* presented to them on the computer screen (their visual representation). The transformation from a visual to an object representation involves the structuring of sensory data into objects, and the grouping together of those objects.

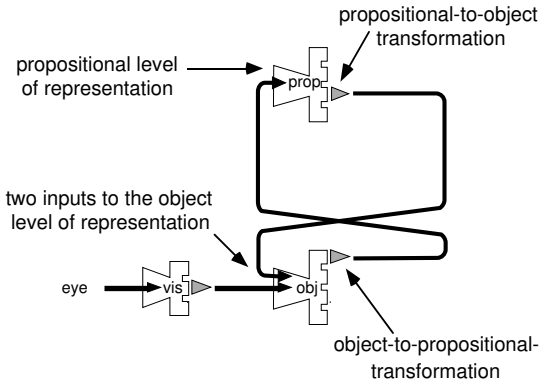

Figure 2.2: the exchange of representations between the object and propositional levels

The visual-to-object transformation is affected by the clarity of the visual representation, so that detailed, high-resolution displays will be easier to convert into object representations than jagged, low resolution displays. It also develops with experience, so that familiar visual patterns can be converted into object structures more accurately than novel patterns - essentially, the more often a representation has been transformed in the past, the easier it becomes to transform in the future. This is one way that experience can affect perception. Another way is shown in Figure 2.2.

The **visual-to-object** transformation process structures and groups the visual scene

Propositional representations

Just as the visual-to-object transformation interprets the visual representation to produce a more abstract, but more structured object representation, the object representation can be interpreted to produce an even more abstract and structured 'propositional' level of representation. This new level contains factual, everyday knowledge about the objects - their names and properties, and the way that they can be expected to relate to each other and to interact. In the same way that the visual-to-object transformation added structural information that wasn't necessarily present in the sensory data, so

the object-to-propositional transformation adds a meaningful identification of objects that isn't necessarily present in the object representation.

For example, a pattern of light and shade can be interpreted by the visual-to-object transformation as a set of flat, square surfaces belonging to two separate but overlapping objects, and the object-to-propositional transformation can then interpret this as a book placed on top of a notepad. It too develops with experience, so that familiar objects in their normal positions are more easily recognised than unfamiliar objects, and than objects in strange orientations. The end result of this sequence of mental processing is that we are able to recognise and identify objects, and to access knowledge about the way they behave and what they do.

The **object-to-propositional** transformation process identifies and relates objects

The remaining part of Figure 2.2 is perhaps the most important. It shows a third transformation process taking place, this time a propositional-to-object transformation. Although all transformations develop with experience, and so allow a slow form of learning to take place, the addition of this third transformation provides a more immediate way for knowledge and expectations about the world to influence perception. It takes the propositional representation, and interprets it to feed a new object representation back to be combined with the visually derived object representation. The object representation that results is therefore really a blend of external data obtained by the senses from the world, and internal data constructed from a mental, propositional representation of the world.

The **propositional-to-object** transformation process feeds back information about object structure

The 'active nature of perception' that we began this guide with is becoming much more active: with the addition of this feedback loop, it becomes possible for the object representation that a viewer forms at one moment to influence the object representation that is formed the next moment. The object representation is receiving information from both the visual-to-object transformation and the propositional-to-object transformation.

The object representation that is perceived is a **blend** of information from visual and propositional sources.

The representation that the object-to-propositional transformation actually uses is a result of these two inputs being blended together: the parts that match reinforce each other, and parts that don't match are discarded. This can be of great benefit in perception, where the sensory, visual level of representation is often incomplete or distorted. When the visual-to-object transformation is unable to produce a clear object representation, the contribution of propositional knowledge allows the viewer's expectations and knowledge about the world to clarify matters. We'll go into more detail about different sorts of blending in Section 7, but for now we will concentrate on its effects at the object level.

The object in Figure 2.3 has the identity 'teddy bear' for a viewer who has already learnt the propositional representation of such an item, and knows that teddy bears generally have a head with ears that stick out, and limbs that are spread out. As soon as the object-to-propositional transformation produces an identification of the shape as a teddy-bear, however weak, the viewer's propositional knowledge of teddy-bears can be brought into play. The propositional-to-object transformation can produce details about what teddy-bears ought to look like. Where these match the visually derived data, the shape can be interpreted as fitting the propositional identification. Slight differences between the propositionally derived representations and the visually derived representations cancel each other out, and do not form part of the object representation. The gaps in the outline, for example, become unimportant. The object-to-propositional transformation now has an even more bear-like representation to work with, and so the feedback between object and propositional levels of representation becomes progressively stronger.

Propositional representations can fill in gaps in object representations derived from incomplete visual representations.

Even if the viewer of this Figure has no idea of what they are about to see, the extreme familiarity of this outline enables the feedback between the object and propositional representations to settle on this interpretation very rapidly, perhaps

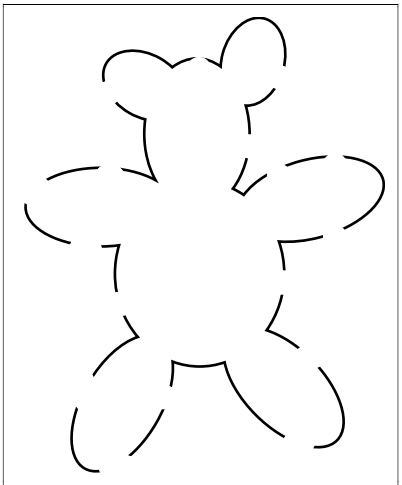

Figure 2.3: propositional knowledge helps to 'complete' the object representation of this figure

ignoring any visual features that did not quite 'fit'. People have a strong tendency to give objects nameable identities if they possibly can, reflecting the extraction of propositional knowledge. These names then affect the way the objects are perceived. The propositional influence on the perception of familiar forms like this is very resistant to distortions in the shape, provided that key invariants between the objects are met. If Figure 2.3 is looked at upside-down, for example, a different teddy bear can be perceived (or perhaps the longer 'ears' now make it look like a rabbit).

For someone who has never seen a teddy bear, and who has no other propositional representation that 'leaps in' to influence the perception of a single object, the form may appear to be a number of overlapping circles and ovals, but the absence of arcs in the centre of the form completing or even continuing these circles means that it is much more likely to be seen as a single, irregular shape.

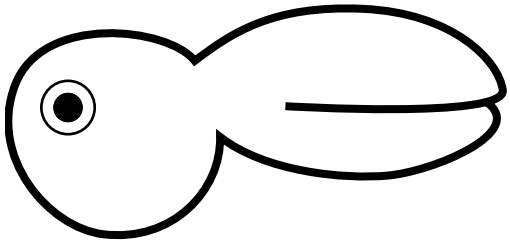

Figure 2.4: ambiguous form

The feedback loop between the propositional and object levels of representation tries to settle on one consistent interpretation of a figure. The form in Figure 2.4 can be seen as either a rabbit (looking to the left) or a bird (looking to the right), but it cannot be seen as both at the same time: the perception must 'reverse' between the two interpretations. Notice that the propositional identity given to the Figure constrains the structure of the object representation – the beak becomes a pair of ears, and the direction that the eye is looking changes. These structural changes in an object representation that has been derived from a single visual representation are indicative of propositional knowledge being brought to bear.

Propositional representations help the object representation settle on one interpretation of ambiguous figures

Once an object has been propositionally identified, we are able to 'go beyond' the available sensory data to use our knowledge about the world to enhance its object representation. If we are told that an object is 'round', or has a 'hole', then we can combine the sensory information that is available now with information that we have experienced in the past as being common to 'round' or 'holed' objects (Figure 2.5). If we were told that it was 'round', we might actually interpret it as 'spherical', even if the appropriate sensory information (such as shading) is not immediately available. If we were told that it was a 'hole', we might be able to perceive some visual features as belonging to another object that is visible through it.

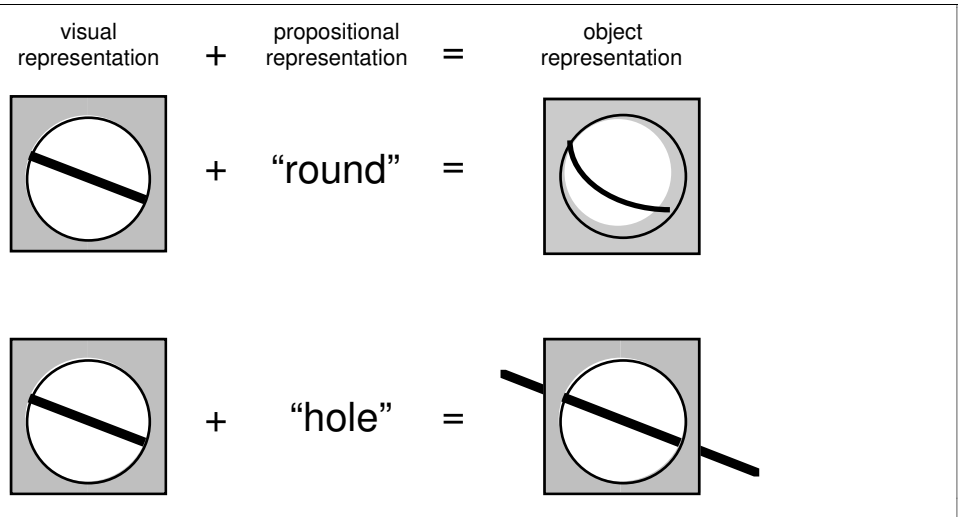

Figure 2.5: the same visual representation can result in different object representations, depending upon the contribution of the propositional representation

The perception of ambiguous figures depends on what the viewer knows, and what they expect to see.

If it were not for the contribution of propositionally derived knowledge about objects, we would be unable to use simple verbal labels like 'book' or 'desk' in the structure diagrams of Section 1. We'd have to use descriptions like 'flattened

cuboid with one side convex along one axis and the opposite side concave and slightly ridged', or 'flat horizontal rectangle with four thin vertical blocks attached underneath at each corner'.

These long-winded descriptions are more like the actual products of the visual-to-object transformation, which can identify shapes and group them together structurally, but cannot identify what they are. Of course, the structure diagrams and transition path diagrams would be impractical to draw if we tried to use such detailed descriptions, and so we use the propositionally-based labels instead, but the examples we presented earlier all contained object representations.

The psychological subjects and the parts of the predicate were all objects, and the point of constructing them was to identify how people would be able to move their focus of attention between objects in the world, whether they were real objects, or representations of objects presented on a computer screen.

The complete cognitive system

The three levels of mental representation that we have described so far are sufficient to deal with the perception of visual objects, but they cannot deal with other important aspects, such as the perception of sound, or the use of language, nor the individual's physical actions. To account for these aspects we will have to add more levels of representation, and more processes.

The complete cognitive system is shown in Figure 2.6, which contains the visual and object levels at the bottom, and the propositional level in the middle. The different levels of representation are now linked together in a 'network' to show that they can exchange information with each other.

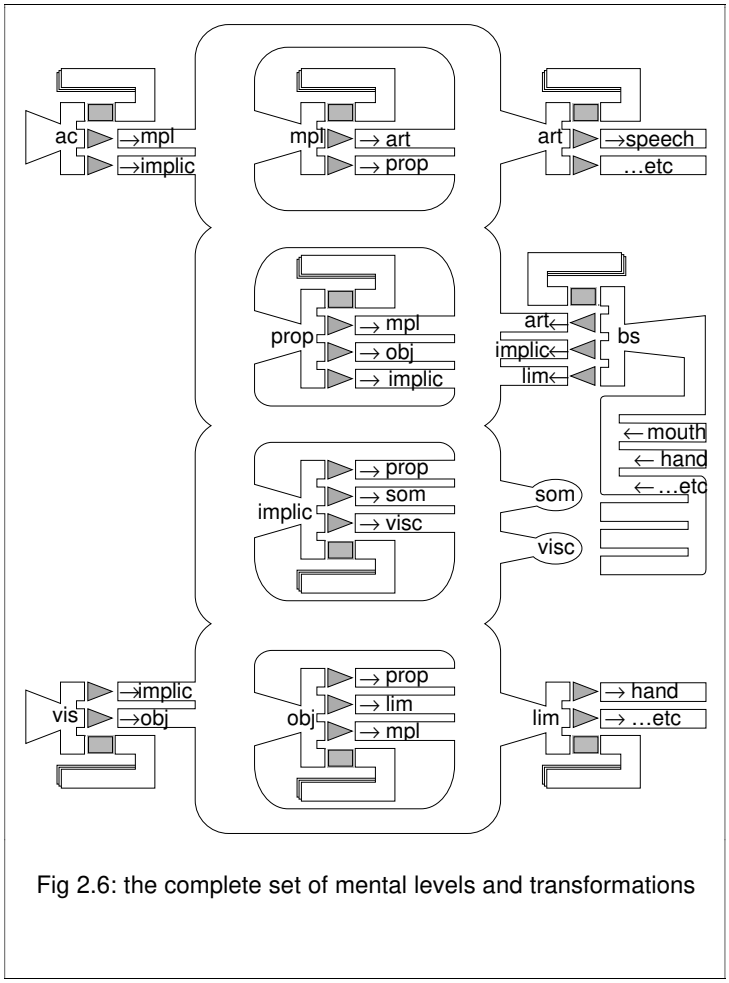

Fig 2.6: the complete set of mental levels and transformations

At the top of the figure are two levels that resemble the visual and object levels, but which deal with acoustic information ('ac') and 'morphonolexical' information ('mpl') respectively. The acoustic level does the same for sensory information from the ears that the visual level does for sensory information from the eyes, and just as the object level is a more abstract, structured representation of visual information, so the 'morphonolexical' level is a more abstract, structured representation of sound. Its name reflects the fact that it contains information about all sorts of sounds, particularly our human speciality, language. It is also crucial in the perception of other structured noises, such as the tones and rhythm of music, as well as the beeps made by computers.

The acoustic and visual levels both encode sensory information, or the 'input' to our minds. On the right of the figure are two levels that encode our mind's 'output', the 'articulatory' and 'limb' levels ('art' and 'lim'). Both these levels represent physical, motor actions that we intend to produce. The articulatory level specialises in controlling the detailed motion of the mouth, lips and tongue required for us to produce sound output such as speech, while the limb level controls other physical actions, such as hand and eye movements.

As their positions in the diagram show, the representations at these levels are mainly produced by transformations of the morphonolexical and object levels, but they also receive information from the 'body state' level ('bs'), which is a third source of sensory information. This information represents all of the touch, smell and taste sensations that our body detects, as well as information from internal sensations such as the position of our arms and legs, and the state of our muscles.

The transformations of this level of representation are important in providing 'feedback' to regulate and co-ordinate our physical actions, because the limb and articulatory levels are representations of intended actions that have been mainly produced from the morphonolexical and object levels, but which have been blended with information from the body state level. If you've ever had a tooth filled, for example, you'll know how difficult it is to rinse your mouth out while the anaesthetic is preventing you using the body state information to detect the position of your lips - even if you are grateful for the absence of pain from your tooth.

The three sensory levels of representation, and the propositional level, can all be used to produce the final level of representation we need, the 'implicational' level ('implic'). This is the most abstract level of all, and it represents the general meaning of information. So if you see something red, the visual-to-implicational transformation produces a representation of all the things that you have learnt red to mean – not just the fact that the object is red. If you are simultaneously hearing a continuous bell ringing, the acoustic-to-implicational transformation will be producing a representation of the general meaning of the sound of bells – not just the fact that you are hearing a bell. The 'facts' are propositional; their meaning is implicational.

When the outputs of these to transformations are blended together to form a single implicational representation, all of the common elements combine, in the same way that the object representations derived from visual and propositional representations could be combined. In the example of 'redness' and 'bells', you would hopefully form the implicational representation of 'dangerousness'. The implicational-to-propositional transformation could then turn this into a direct, propositional fact that there is something dangerous around.

Figure 2.6 also shows transformation processes that turn the implicational representations into physical effects within our body, and so can affect our moods ('som' means somatic, and 'visc' means visceral). The way we interpret the world can interact with the way we feel, which can in turn affect the way that we interpret the world. When you are feeling tired and stressed, working to complete a piece of work before a deadline, your computer's 'beep' can seem very much more annoying than when you are feeling fresh and alert.

Nine subsystems

Figure 2.6 includes many more transformations than we have described so far – all of the shaded triangles indicate a possible transformation from one level of mental representation to another. To keep the diagram simple, we have not included all of the arrows that link the different levels together. The object level of representation, for example, can be used for three different transformations. As well as the object-to-propositional transformation that we have already described, there is the object-to-limb transformation that controls motor actions, and an object-to-morphonolexical transformation, which develops as we learn to read fluently, and enables us to 'hear' words in our mind as we look at text.

For convenience we can think of all of the transformations that can be made from a given level of representation as part of the same cognitive subsystem, there being one subsystem corresponding to each level of mental representation. Each of these subsystems receives the representations it is specialised for, possibly from a variety of sources, and can produce a variety of representations for other subsystems to receive.

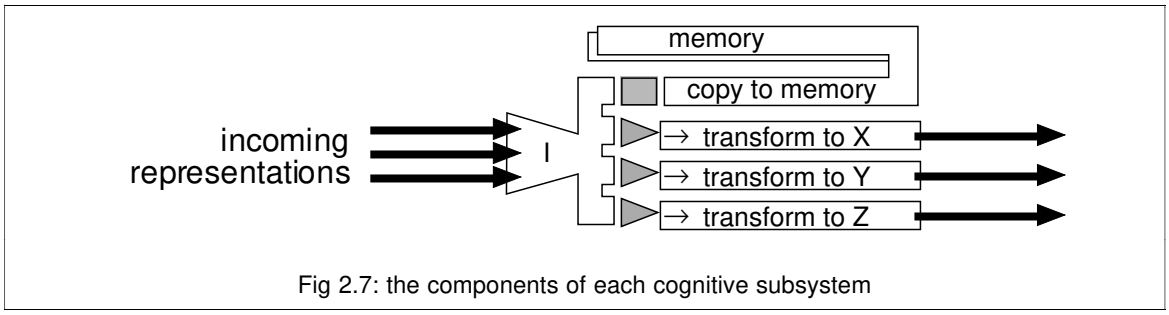

Fig 2.7: the components of each cognitive subsystem

A schematic picture of a typical subsystem is shown in Figure 2.7. It receives representations from the left, and each of the shaded triangles indicates a different transformation process, with the transformed representations passing out to the right. The shaded rectangle indicates an additional process, but instead of transforming the incoming representations, this process copies them, unaltered, into memory.

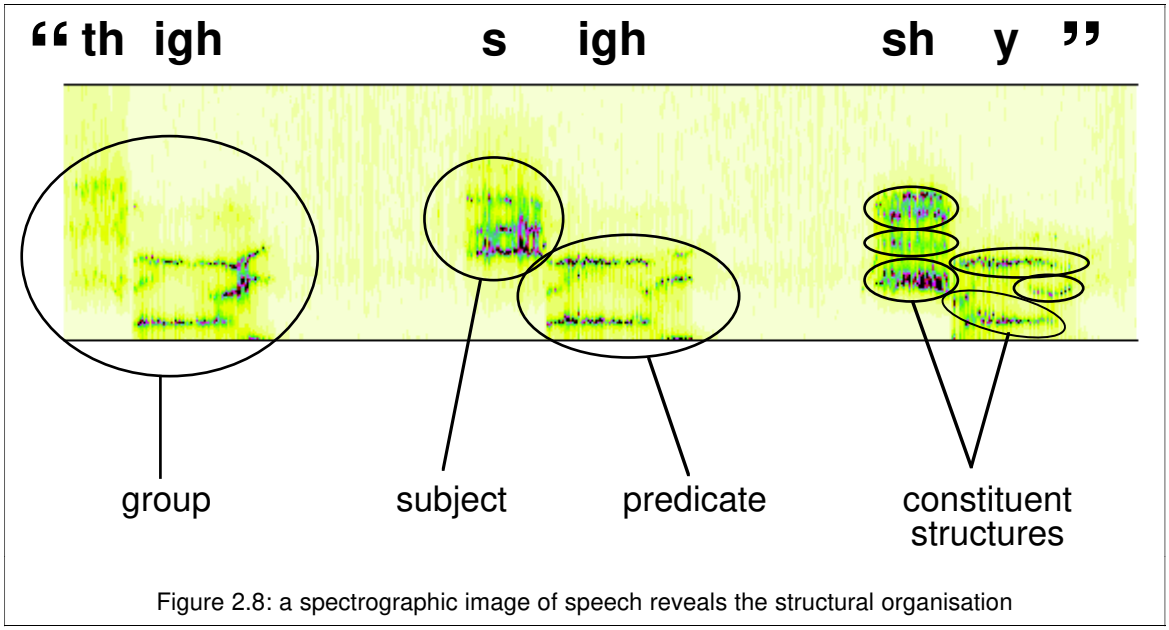

Figure 2.8: a spectrographic image of speech reveals the structural organisation

Each level of mental representation is processed by a different **cognitive subsystem**.

As you can see from Figure 2.6, each of the nine individual subsystems has such a copy process, and each has its own memory. These memories allow each subsystem to learn about the representations that they receive, so that if a representation ‘looks like’ something that has been received before, it can revive a representation in memory. The revived representation can then be used by one of the transformation processes instead of the incoming representation.

All nine subsystems have a **common architecture**, including their own memory.

So if your object subsystem receives a representation of ‘ball’ from the propositional subsystem, you can produce a mental image of all sorts of balls from memory, without having to use any visual representations at all. Similarly, the propositional subsystem can produce a morphonolexical representation of the word ‘ball’, and you can imagine the word ball being spoken in many different voices - but normally, you would hear it in your own voice, since that is the voice that you hear most often, and so your morphonolexical memory is mainly filled with it.

This example makes the point that the same propositional representation, the concept of a ball, can be used to revive quite different sets of memories, depending upon the level of representation that it is transformed into, and which subsystem’s memory is subsequently accessed. Since the subsystems process different levels of representation, their memories contain different types of information.

The **content** of a memory depends upon the subsystem in which it is stored.

While each of the levels of representation contains a different sort of information, all representations are structured in the same general way as the object representations, and can all be thought of as having a group, subject-predicate and constituent structure. The spectrographic image of the energy at different frequencies within a fragment of speech shown in Figure 2.8 reveals this structure in an acoustic representation. The generic structure of information, regardless of the level of representation, is one of the consequences of the common architecture of the subsystems. It means that the same techniques of structural analysis, using structure diagrams and transition path diagrams, can be used to understand cognition whatever the level of representation.

All levels of representation can be described as having groups, subject-predicate and constituent structures. Structure diagrams and Transition Path diagrams can be used for all levels of representation.

Interacting Cognitive Subsystems

Each of the nine subsystems is continually receiving representations, copying them into its own memory, and transforming them into other representations, and they all act in parallel with each other. The nature of the information that each subsystem processes is summarised in Figure 2.9. As the examples we have used in this section have emphasised, it is the combination of representations from different sources, and the exchange of representations between levels of representation, that provides human cognition with its elaborate and complex richness.

The framework of cognition that we have described is called 'Interacting Cognitive Subsystems', or ICS, since it is the interaction between the subsystems, rather than their individual action, that is seen as most important in understanding the way that we perceive, think and act.

Sensory subsystems:		
	Acoustic:	sound frequency (pitch), timbre, intensity *what we hear in the world*
	Visual	light wavelength (hue), brightness, saturation *what we see in the world*
	Body State	stimulation type, location, intensity *what we feel in our body*
Perceptual subsystems:		
	Morphonolexical	Abstract structure of sounds, especially speech *what we hear in our head, our mental voice*
	Object	Abstract structure of visual objects, their position and motion *what we see in our head, our mind's eye*
Central subsystems:		
	Propositional	the identities of objects, their relationships, and facts about them *what we know as facts about the world*
	Implicational	ideas about the real 'meanings' of events, situations and emotions *what we know as 'feelings' or 'impressions'*
Effector subsystems:		
	Articulatory	force, target positions and timings of vocal and gestural muscles *what we intend to say, subvocal speech and gesture*
	Limb	force, target positions and timings of skeletal muscles *what we intend to do, physically.*

Figure 2.9: the nine cognitive subsystems, the types of information that they each deal with, and their subjective qualities (italics).

To understand how people will perceive, learn about, and use an interface design, it helps to think about all of the sources of information that they will be using as their cognitive processes operate. The interactions between the subsystems mean that all of the perceptual and central subsystems are influenced by more than one source of information. The implicational level is built up from a blend of transformations from sensory inputs and the current propositional representation. The propositional representation takes some input from the implicational level, some from the structure of the visual information, and some from the structure of the sound information. Sound is structured according partly to aspects of the raw sensory data, partly from the propositionally-based expectation about what we are hearing, and also partly from the interpretation of the object level, if we happen to be looking at text or other linguistically related information (including lip movements). Similarly, understanding object representations means considering the visual details as well as the contribution of the propositional subsystem.

All perceptual and central subsystems receive representations from more than one source.

In the following sections we will look in more detail at some of the ways that visual information can be used to add structure to an object representation, before moving on to look at the way that propositional information can be blended. This will introduce us to blending at other levels of representation.