Department of Computer Science,
University of Otago

Technical Report OUCS-2017-04

Sensorimotor cognition and natural language syntax Part II

Author:

Alistair Knott
Department of Computer Science, University of Otago, New Zealand

Department of Computer Science,
University of Otago, PO Box 56, Dunedin, Otago, New Zealand

http://www.cs.otago.ac.nz/research/techreports.php

Sensorimotor cognition and natural language syntax
Part II[1]

Alistair Knott, Dept of Computer Science, University of Otago
`alik@cs.otago.ac.nz`

January 10, 2017

---

[1] This manuscript is still in draft form—comments welcome!

Contents

18 A new model of LTM, and its connection to the sensorimotor system 517

19 Relationship of the theory to existing work 536

Chapter 9

Introduction

This book is a continuation of the idea I developed in my earlier book, 'Sensorimotor Cognition and Natural Language Syntax' (Knott, 2010). In that book, I suggested that the syntactic structure of a sentence reporting a concrete episode could be interpreted as a description of sensorimotor processing. I expressed this idea using the syntactic framework of Minimalism (Chomsky, 1995), in which every sentence has two syntactic representations: a **phonetic form** (**PF**) and an underlying **logical form** (**LF**). My proposal was that the LF of a sentence S reporting a concrete episode E can be characterised as a description of the sensorimotor processes involved in actually experiencing the episode E. In the earlier book, I focussed on a single syntactic construction (a transitive clause) when presenting and motivating this proposal. Obviously I must consider a wider range of constructions. In the current book I examine how the original proposal extends to other syntactic constructions.

Chapter 10

A sensorimotor characterisation of noun phrase syntax

10.1 Introduction

In the book so far we have focussed on ‘proposition-sized’ units: the sensorimotor model has been about a reach-to-grasp *episode*, and the corresponding syntactic model has been about the structure of a transitive *clause.* In the next two chapters, I will consider units below the level of propositions: the sensorimotor focus will move to objects and attentional routines, and the syntactic focus will move to noun phrases—or in current syntactic parlance, **determiner phrases** (**DPs**). The aim is to extend the sensorimotor interpretation of syntactic structure proposed in Chapter 5 to DPs.

Obviously, perceiving a concrete episode involves perceiving the objects which participate in it, and representing an episode involves representing these objects. The sensorimotor model of episodes given in Chapters 2 and 3 already includes some suggestions about how objects are perceived and represented, but it does not give much detail. Likewise the syntactic model given in Chapter 4 says something about where DPs can appear in a transitive clause, but does not go into detail about the syntax of DPs. In the next two chapters I will expand the syntactic model of DPs and the sensorimotor model of object perception, and ty to link these two models together.

The division of labour in the next two chapters roughly reflects the fact that DPs can contribute to clauses in two quite different ways. The most obvious thing a DP can do is to refer to an object. The two DPs in our example sentence both have this function:

(10.1) *The man* grabbed *a cup.*

The most obvious sensorimotor interpretatation of the internal syntax of a referential DP is as a trace of the sensorimotor mechanisms involved in delivering an object representation. This is in fact the main idea I will pursue in the current chapter. But there are many DPs which cannot be understood referentially. The most important class of these are **quantified DPs**, examples of which are given below.

(10.2) *Most men* grabbed *a cup*.

(10.3) *No man* grabbed *a cup*.

None of the DPs in these examples straightforwardly ‘refer to objects’. (The DP *most men* picks out a different set of men depending on what is predicated of it; *a cup* may pick out a different cup for each man, and *no man* obviously has no referent at all.) A general ‘sensorimotor’ interpretation of DP syntax will have to cover quantified DPs as well as referential ones. In fact, what syntacticians want is a unified account of referential and quantified DPs, which explains why DPs can function in both ways. I will focus on referential DPs in this chapter, but I will move towards a unified account in Chapter 11.

In this chapter, I will look for a sensorimotor interpretation of the internal syntactic structure of DPs, drawing on a more detailed model of the perception of objects and their representation in working memory. In Minimalism, the syntactic structure of a DP is strongly right-branching, just like that of a full clause. A preview of the LF of the DP *the man* is given in Figure 10.1. The highest projection DP is headed by the determiner *the*. This introduces a functional projection called **NumP**, which contributes the ‘grammatical number’ of the DP—in this case, singular. NumP in turn introduces the familiar NP projection. Without going into details, it is clear that the general interpretation of LF proposed in Chapter 5 makes some strong predictions about the structure of the sensorimotor processes involved in delivering an object representation. In particular, it predicts that these processes will be organised into three sequential stages. In this chapter, I will argue that this prediction is correct.

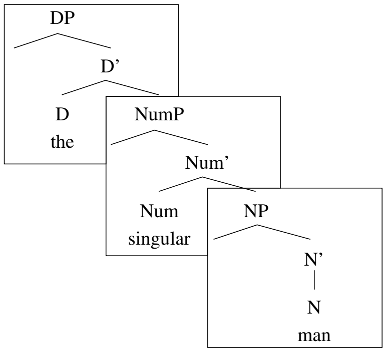


Figure 10.1: The LF structure of a determiner phrase: preview

I begin in Section 10.2 by introducing a basic model of DP syntax, which omits the NumP projection. In Section 10.3, I present an initial sensorimotor interpretation of this basic model, which draws on the account of object perception given in Chapter 2. This initial interpretation fits well with the general sensorimotor interpretation of LF proposed in Chapter 5. But both the syntactic and the perceptual accounts need to be extended. The model of DP syntax must be revised to separate out the notion of grammatical number from the notion of reference. A revised model, introducing the NumP projection, is given in Section 10.5. The model of object perception must be extended to cover attention to and categorisation of *groups* of objects, as well as working memory representations of objects and groups. An extended model of object perception is given in Section 10.5, and a model of working memory for objects and groups is given in Section 10.6. In Section 10.7, I present a more detailed sensorimotor interpretation of the syntax of referential DPs.

10.2 A simple syntactic model of DPs

Until the mid-eighties, it was assumed that referential or quantifying phrases like *the hat* or *every hat* were headed by nouns, as shown in Figure 10.2(a). This idea was forcefully challenged by several linguists, most prominently Szabolcsi (1987) and Abney (1987). Abney proposed an alternative model, in which such phrases are basically **determiner phrases** (DPs), which introduce noun phrases as complements, as shown in Figure 10.2(b). The 'DP hypothesis' can be motivated on syntactic grounds, but it also fits well with semantic accounts of reference and quantification which were being developed around the same time, and which are now widely accepted. In this section I will summarise some syntactic arguments for DP, and then briefly introduce a semantic model which fits with it. For more details in each case, see Alexiadou *et al.* (2007).

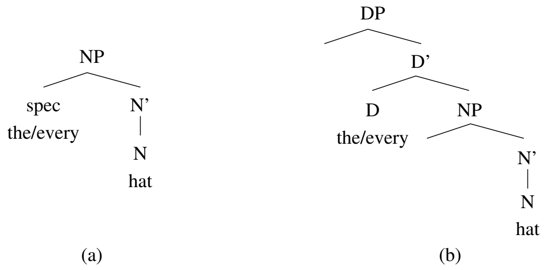


Figure 10.2: Syntactic analyses of *the hat*: (a) as an NP; (b) as a DP

10.2.1 Abney's original argument for the DP hypothesis

Abney begins by noting that there are a surprising number of similarities between noun phrases and full clauses. For instance, the concepts of 'subject' and 'object' which operate in clauses seem to have correlates in possessive noun phrases:

(10.4) *Peter* has *a hat*

(10.5) *Peter*'s *hat*

In addition, the clause-level phenomenon of agreement between the subject and its verb seems to have noun-phrase-level correlates in several languages. For instance, in Hungarian, the noun in a clause shows agreement with the possessive subject-like element (c.f. Szabolcsi, 1987):

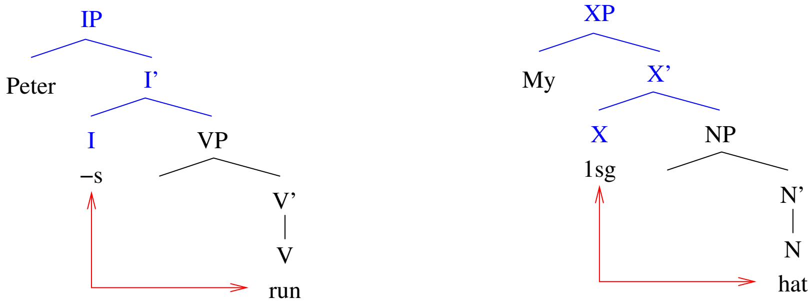
	az	en	kalap-om
(10.6)	the	I:NOM	hat-1sg
	‘My hat’		

	a	te	kalap-od
(10.7)	the	You:NOM	hat-2sg
	‘Your hat’		

	a	Peter	kalap-ja
(10.8)	the	Peter:NOM	hat-3sg
	‘Peter’s hat’		

The important thing to note about these examples is that the noun *kalap* is inflected differently according to the person and number features of the possessor, in a way reminiscent of subject-verb agreement (c.f. e.g. *I am*, *you are*, *Peter is*).

In the model of clause-level subject-verb agreement given in Chapter 4, the inflection on the verb ‘belongs’ at the head of its own functional projection, IP, as shown in Figure 10.3(a). This projection introduces the subject, which explains why the inflection agrees with the subject. To explain how the inflection ends up on the verb, some form of head movement is stipulated: in GB the inflection lowers to the verb, and in Minimalism the inflected verb raises to I to ‘check’ its inflection. If this account of agreement is correct, this suggests a similar account of agreement in possessive nominals, in which the noun’s inflection ‘belongs’ at the head of a higher functional projection introducing the possessor, as shown in Figure 10.3(b).



(a) (b)

Figure 10.3: (a) Subject-verb agreement in the sentence *Peter run-s*; (b) Possessor-noun agreement in the nominal *My hat-1sg*

Abney argues that the analysis proposed for possessive nominals in Figure 10.3(b) can also serve for ‘regular’ nominals like *the hat* and *every hat*. The higher functional projection XP can be interpreted as the projection created by the determiner—i.e. as a DP—as shown in Figure 10.2(b). In X-bar theory, every word projects its own XP, containing slots for a specifier and a complement; the theory’s appeal rests on its ability to give useful definitions of specifier and complement which generalise across syntactic categories. Abney's revised analyses of nominal constructions as DPs consolidate the generalisations captured by X-bar theory, and capture important generalisations between nominal and clausal constructions.

10.2.2 Some further syntactic arguments for N-to-D movement

In Section 10.2.1 I appealed to the idea of movement between N and D heads in Hungarian possessive phrases (see Figure 10.3(b)). But in order to properly motivate the idea of N-to-D movement, it is important to show it has useful application beyond this one construction. In this section I will discuss two independent constructions where the concept can be usefully applied.

I will begin by considering another case where there is a strong analogy between DPs and clauses. As discussed in Section 10.2.1, possessive nominal constructions have very obvious clausal counterparts; for instance in *Dan's love of himself*, there is a subject-like element (*Dan*), a verb-like element (*love*) and an object-like element (*himself*). In Hebrew, possessives can be expressed in a nominal construction called a **construct state**, where the verb-like element appears first, as shown below.

ahavat dan et acmo
(10.9) love Dan ACC himself
'Dan's love of himself'

The surface order of elements here is 'verb subject object'. Recall from Section 4.7 that in a transitive clause, the subject and object originate at the specifier and complement of VP, and VSO word order results from raising the head of VP to the head of the higher IP projection, as shown in Figure 10.4(a). Ritter (1991) has influentially argued that the Hebrew construct state derives from a similar structure, where the subject-like and object-like elements appear at the specifier and complement of NP, and the verb-like head of NP (*ahavat*) raises to the head of a higher DP projection, as shown in Figure 10.4(b). The positions of the subject-like and object-like elements within a single XP are motivated by the fact that the object can be a reflexive pronoun—as it is in our example. (Reflexive

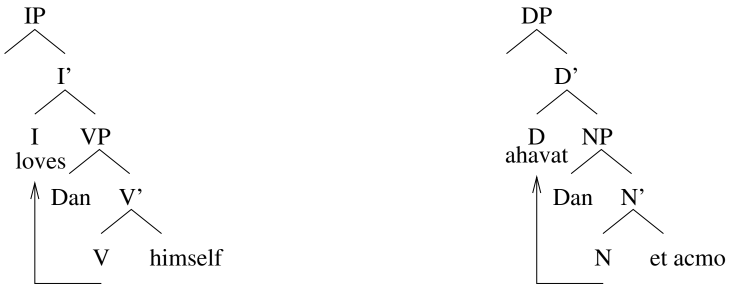


Figure 10.4: (a) Derivation of VSO order in a clause. (b) Derivation of 'VSO-like' order in the Hebrew construct state

pronouns and their antecedents can only appear in certain quite local structural configurations, according to a component of syntactic theory called ‘binding theory’, which I will not discuss here.) The ‘natural’ position of the verb-like element *ahavat* is at the head of this XP, in between subject and object, because this is the position which assigns subject and object thematic roles. So in a construct state, there is a strong indication that it has moved out of its original position to a higher functional head. The suggestion that it has moved to the head of DP comes from the fact that the regular Hebrew definite determiner (*ha-*) cannot appear in the construct state:

(10.10)	*	ha-ahavat	dan	et	acmo
		the-love	Dan	ACC	himself

Ha- normally appears at the head of DP—but of course this position can only be occupied by a single lexical item. Arguing that the noun *ahavat* has moved into the empty D in the construct state allows us to explain why *ha-* cannot appear in this construction. Ritter suggests that a phonologically empty item at D functions to convey the possessive character of the construction—specifically, she suggests that it assigns **genitive case** to the subject, in the same way that that an overt particle like *of* can indicate the possessor in English.

While Ritter’s argument for N-to-D movement appeals to the idea that D cannot be occupied by two items simultaneously, there are other arguments which appeal to the idea that D needs to be occupied by at least *one* item. One such argument comes from an analysis of Italian proper nouns (Longobardi, 1994). In Italian, a proper noun which is modified, for instance by a possessive, is standardly introduced by a definite determiner:

	Il	mio	Gianni
(10.11)	The	my	Gianni
	‘My Gianni’		

In fact, modified proper nouns can also appear without determiners. But in these cases, the proper noun must appear to the left of the modifier:

(10.12)	Gianni	mio
	Gianni	my

(10.13)	*	Mio	Gianni
		My	Gianni

Longobardi argues that when there is no explicit determiner, the proper noun *Gianni* must raise to fill the empty D position, creating the structure illustrated in Example 10.12. If it does not, as in Example 10.13, the construction is ill-formed.

Another kind of argument for N-to-D movement comes from the fact that in some languages, determiners can be realised as inflections (‘clitics’) on nouns. For instance, in Romanian, a demonstrative is realised as an independent word positioned to the left of adjectives and the noun (e.g. Example 10.14), but a definite determiner is realised as an inflection on the noun, which then appears to the left of adjectives (Example 10.15).

(10.14) Acest sàrac bàiat
This poor boy
'This poor boy'

(10.15) Bàiat-ul sàrac
Boy-the poor
'The poor boy'

Grosu (1988) proposes that in Example 10.15, the inflection *-ul* appears at D, and the noun *bàiat* raises to this position to provide a stem for it to attach to. This explains why the noun appears to the left of adjectives when it takes a clitic determiner.[1] Clitic determiners are also common in Scandinavian languages, where a similar N-to-D raising analysis can be given (see e.g. Taraldsen, 1990).

In summary, the DP hypothesis, along with the assumption that N heads can raise to D, permits an elegant analysis of a range of nominal constructions in different languages. I would not want to claim that it applies easily to *all* languages. There are many languages where determiners appear postnominally as independent words, often at some distance from the noun; see e.g. Haitian creole (Lumsden, 1989) and West African languages like Fongbe and Yoruba (Roberts, 2007). Postnominal determiners are easy to account for on the DP hypothesis if we can allow the DP head to appear after its complement—but in our interpretation of X-bar theory, heads strictly precede their complements. For now, I will leave postnominal determiners as a topic for future discussion.[2] Aside from these cases, the assumption that N can raise to D is very productive for a wide variety of languages.

[I think I need to put something in here about adjectives: e.g. size and colour adjectives (*The big brown dog*/*Le grand chien brun*).]

10.2.3 The semantic contribution of determiners: a model of reference in discourse

Semanticists have quite well-developed ideas about what determiners and nouns contribute to the meaning of a sentence. In this section I will briefly introduce these ideas, using a model of reference and quantification which was also introduced in the eighties, by Kamp (1981; see also Kamp and Reyle, 1993) and Heim (1982). I will use Kamp's term 'discourse representation theory' (**DRT**) for the model. DRT is important, because it gives a unified semantic treatment of referential and quantified DPs—and moreover, provides the basis for an account of the difference between definite and indefinite DPs (e.g. between *the dog* and *a dog*).

---

[1] In a Minimalist analysis, the noun would be generated with its determiner inflection at N, and would raise to D to 'check' this inflection. This presumably allows the option of inflected nouns raising 'covertly' to D, but appearing to the right of adjectives in surface sentences.

[2] It is worth noting that in the languages just mentioned, determiners appear to contribute syntactically to their host clause as well as to nominal constructions, which perhaps places them outside the scope of a basic account of DP-internal syntax.

The basic proposal in DRT is that the semantic content of a sentence should be modelled not as an independent proposition, but as a contribution to a representation called 'the discourse', which grows and evolves as succcessive sentences are interpreted. The discourse is formalised as a **discourse representation structure** (**DRS**), which consists of a set of **referents**, which denote objects in the world, and a set of **conditions**, which specify facts about these objects. A DRS is represented diagrammatically as a box, with one compartment for referents and one for conditions, as shown in Figure 10.5.

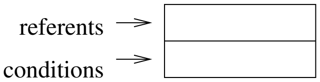


Figure 10.5: A discourse representation structure (DRS)

In DRT, sentences can introduce new referents into the discourse, but can also refer to referents already introduced. These functions are carried out by different types of DP. **Indefinite DPs** (e.g. *a dog*) introduce new referents, while **definite DPs** (e.g. *the dog*) refer back to existing ones. To take an example, consider the following two-sentence discourse:

(10.16) *A dog* came in. *The dog* sat down.

Before the discourse is interpreted, we begin with an empty **working DRS**, as depicted in Figure 10.6(a). Interpreting the first sentence (*A dog came in*) using this working DRS

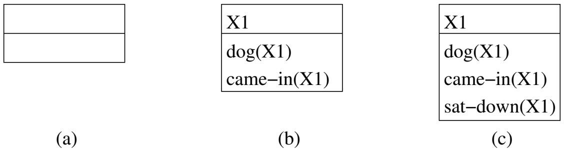


Figure 10.6: A series of updates to the 'working DRS' as Example 10.16 is interpreted

results in the DRS shown in Figure 10.6(b), containing a single referent (X1), and two propositions about this referent. The first proposition, dog(X1), is contributed by the indefinite DP *a dog*. The second, came-in(X1), is contributed by the sentence. The DRS in Figure 10.6(b) now becomes the new working DRS, in which the second sentence (*The dog sat down*) is interpreted. This sentence does not introduce any new referents itself: instead, its definite DP *the dog* 'picks out' a referent introduced by the previous sentence. This is a referent which is already present in the current working DRS (Figure 10.6(b)). This form of reference is termed **presupposition**: the definite DP *the dog* requires there to be a dog in the current working DRS. (More precisely, it requires there to be exactly one referent X in this DRS of which dog(X) is predicated.) If such an object is not found, the sentence containing this DP simply cannot be interpreted—at least, not without additional work. If it is found, as in our example discourse, the property asserted by the sentence is predicated of the retrieved object, resulting in the DRS in Figure 10.6(c).

In DRT, there are various different ways in which nominal expressions can obtain referents. Indefinite DPs create new referents; definite DPs presuppose referents with particular properties. Notice that the element in a DP which indicates how the referent is obtained is the determiner (in the above example *a* or *the*), not the noun. The determiner is what specifies the 'structure' of the semantics of a DP; the noun just fills a slot in this structure. If we want the syntactic contribution of words to mirror their semantic contribution, it makes sense to see determiners creating syntactic positions for nouns, rather than vice versa. In other words, we prefer to analyse referential nominals as DPs introducing NPs (as in Figure 10.2(b)) rather than as simple NPs (as in Figure 10.2(a)). The DP hypothesis is a natural fit for DRT's account of reference.

The notion of a referent is also employed in DRT's account of quantification. A quantifying DP like *every man* introduces a referent too—but it is a referent which is 'bound' by a quantifier, rather than free in the discourse. The notion of quantifiers introducing bound variables is familiar from first-order logic. For instance, the sentence *Every man walks* would be expressed in first-order logic as $\forall x[man(x) \rightarrow walks(x)]$; the variable x is introduced by the quantifier $\forall$, and the uses of this variable within the square brackets are understood as being 'within the scope' of this quantifier. In DRT, a similar notion of bound variables is used to represent quantified propositions, with the innovation that these variables are structurally similar to discourse referents: they can be referred back to presuppositionally, in certain specified contexts. We will discuss DRT's model of quantification in Chapter 11. The important point to note for the moment is that even in quantified DPs, it is the determiner (e.g. *all*, *most*) which introduces the ('bound') referent.

10.3 An initial sensorimotor interpretation of referential DPs

In this section, I will propose a simple sensorimotor interpretation of DP syntax. As already mentioned, I will focus on referential DPs in this chapter. The idea I want to propose is that the syntactic structure of a DP which refers to a concrete object can be read as a description of the processes involved in perceiving that object.

The model of visual object perception presented in Chapter 2 can already serve as the basis for a simple sensorimotor interpretation of DPs. To recap from Chapter 2: there is a distinction between an attentional system, which represents regions of interest in the visual field, and an object classification system, which determines the type of objects presented to it.[3] In Section 2.4 I proposed that the attentional system computes a 'saliency map', principally in the lateral intraparietal cortex (LIP) and frontal eye fields (FEF). The most salient region in the saliency map functions to gate the input to the classification system

---

[3] Maybe refer to Kravitz *et al.* (2008) who questions whether object classification really is invariant to retinal location...

in inferotemporal cortex (IT), so that it only classifies a single region (or perhaps a small number of regions) at any given time. The interaction between the visual pathways for attention and object classification is shown in Figure 10.7, reproduced from Section 2.4.3.

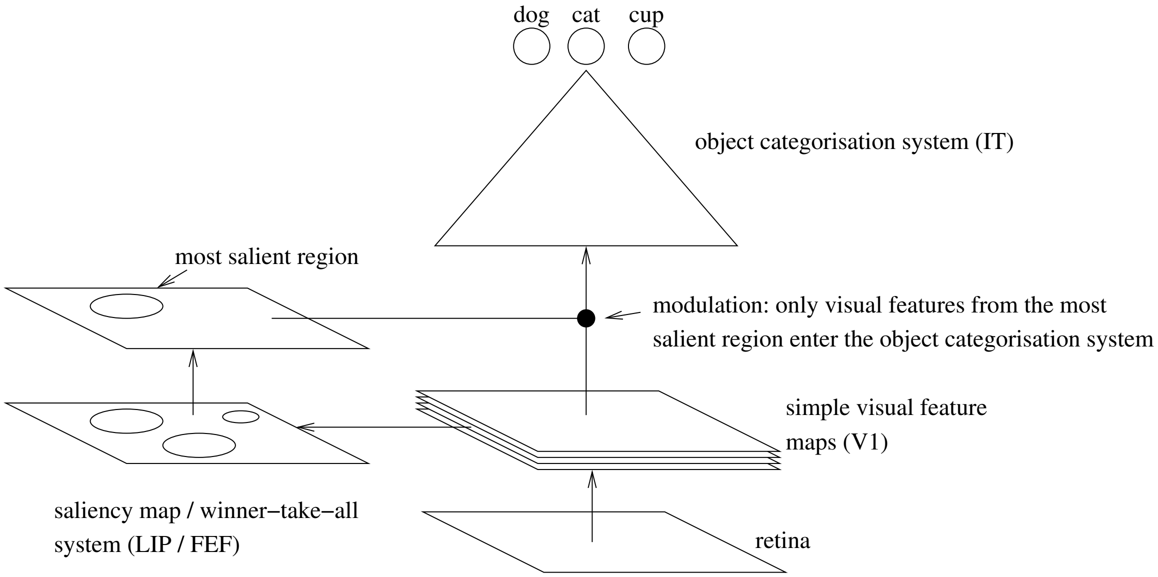


Figure 10.7: The visual pathways for attention and object classification

Recall that the dependence of object classification on attention imposes an ordering constraint on object perception, which was stated as Ordering Constraint 1: a salient region must be selected in the saliency map before the classifier can deliver a category. So according to this model, perceiving an object involves a short sequence of processes: first a region in the saliency map is selected, and then the object occupying that region is classified.

Taking this basic model of object perception as a starting point, I will propose an initial sensorimotor interpretation of the syntax of referential DPs, which adds to the proposals about clause syntax made in Chapter 5.

Principle 10 *The LF of a referential DP denoting a concrete object is a description of the processes involved in perceiving that object.*

Before I motivate this proposal in detail, it is important to note that it makes some intuitive sense: if a linguistic signal is understood as evoking the kind of perceptual routine through which objects are established in the actual world, it is possible to see how the signal can ‘refer’ to an object. (Note that the proposal does not say that referential DPs directly describe objects: rather they describe the cognitive processes through which objects are established. This level of indirection will prove useful when generalising the interpretation of DPs to non-referential cases.) But of course a detailed justification of the proposal must make reference to the details of a syntactic model of DPs and of a model of object perception.

I begin by noting a point which has often been noted before: the decomposition of information in a referential DP seems to mirror the way information is decomposed in object perception. In the syntactic account of DPs just given, the determiner (D) contributes a referent to the discourse, and the noun (N) predicates a type of this referent. These two functions correspond quite naturally to the two components of the perceptual model. Given this parallel, it is quite natural to hypothesise that the determiner describes the operation of selecting a salient region in the saliency map, and the noun describes the operation of activating a type in the object classification system. This is what I will argue.

Principle 11 *At LF, the DP projection of a concrete referential DP describes the selection of a salient region in the saliency map, while the NP projection describes the activation of a type in the object categorisation system.*

As I just mentioned, the suggestion that the architecture of the object categorisation system has reflections in language is not new. For instance, Hurford (2003) argues that the 'predicate-argument' form of the semantic representations which are delivered by language reflects the fact that the perceptual system is decomposed into an attentional system for identifying referents and a classification system for computing their properties. However, Hurford stops short of claiming that the organisation of the perceptual system has reflections in the syntax of language. In Principles 10 and 11 I take this extra step, and interpret elements of syntactic structure as descriptions of perceptual operations.

I suggest that that the extra step is justified becuse the internal syntactic structure of a DP can be very neatly interpreted as a description of processing in the object perception model, given the sensorimotor interpretation of syntax proposed in Chapter 5. As discussed in Section 10.2, the LF of a referential DP consists of a right-branching structure of two XPs: a DP followed by an NP. In Chapter 5 I made the general proposal that a right-branching structure of XPs describes a *sequence* of sensorimotor operations (Principle 3). In our model of object perception, perceiving an object does indeed involve a sequence of operations: we must first activate a region in the saliency map, and then activate a type in the classification system (Ordering Constraint 1). If we assume the sensorimotor denotations of DP and NP just proposed in Principle 11, then their right-branching syntactic arrangement with NP as a complement of DP is exactly what is predicted from our model of object perception and our general sensorimotor interpretation of LF structures.

This is quite a striking result. What we have are two completely separate ways of arriving at the 'DP hypothesis'—the proposal that DPs introduce NPs as complements. One route is via purely syntactic argumentation: the argumentation reviewed in Section 10.2. The other is via a model of object perception (which has nothing to do with syntax at all, but is well motivated in its own right), and a general sensorimotor interpretation of LF syntax (which so far has only been substantiated for a particular kind of transitive clause). The fact that this general sensorimotor interpretation of LF does useful work in another area of syntax can be thought of as a piece of evidence in its favour. It certainly suggests that it is worth looking for reflexes of the perceptual system in the syntax of DPs, as well as just in their predicate-argument semantics.

To sum up so far: there is some suggestion that the LF structure of a (referential) DP can be understood as a description or trace of the processes involved in perceiving an object. However, the stories I have told so far about object perception and about DPs are both rather simple. Both perceptual and syntactic models need to be extended in many ways. In the remainder of this chapter, I will discuss a number of extensions of each model. The question, of course, is whether these extensions allow the above neat perceptual interpretation of DP syntax to be maintained—i.e. whether the more elaborate structure of a DP in the extended syntactic model can still be given a perceptual interpretation within the extended perceptual model. I will argue that they can.

10.4 An extended model of DP syntax: the NumP projection

As discussed in Section 10.2, the DP hypothesis highlights some interesting structural similarities between noun phrases and full clauses. In a clause, the key lexical projection (VP) is introduced by a pair of functional projections (IP and AgrP). According to the DP hypothesis, the lexical projection NP is also introduced by a functional projection—namely DP. In a clause, the head of VP can raise to the head of higher functional projections. According to the DP hypothesis, the head of NP can raise in the same kind of way. In this section, I will outline an argument that there is a second functional projection in the DP, whose head contributes information about ‘grammatical number’—i.e. singular or plural. This argument gives DPs a slightly more complex structure, and makes their structural resemblence to clauses even stronger. The new functional projection is called **NumP**; its role is to contribute the **grammatical number** of a nominal construction. NumP sits in between DP and NP, in the way that AgrP sits in between IP and VP in a clause. The NumP hytpothesis was originally proposed by Ritter (1991), in an extension of her account of DP given in Section 10.2.2. I will summarise her original argument below. But before I do, I will sketch an argument for NumP which is based on a model of DP structure by Zamparelli (2000).[4]

---

[4] Zamparelli’s terminology for nominal projections is a little different. He calls DP ‘SDP’ and NumP ‘PDP’. He also introduces another intermediary projection in between NumP and NP called ‘KIP’, which is used to deal with phenomena we will not consider here.

10.4.1 NumP: a projection introducing grammatical number

Zamparelli's model of DP structure focusses on the role of DPs in two types of clause which we have not yet considered. One of these is a **predicative** clause. Two examples of predicative clauses are given below.

(10.17) This dog is big.

(10.18) This dog is *a collie*.

A predicative clause predicates a property of an individual. The property can be denoted by an adjective, as in Example 10.17; but it can also be denoted by a nominal construction, as in Example 10.18. In this example, the nominal phrase *a collie* behaves very much like an adjective: it contributes a property, which the predicative construction applies to the subject. Such nominals are called **predicate nominals**.

Note that a predicate nominal does not introduce a referent. There is only one referent in Example 10.18, and that is contributed by the subject. If the semantic function of a DP projection is to introduce a referent, as argued in Section 10.2.3, then predicate nominals should not contain DP projections at all. At the same time, predicate nominals certainly can have a determiner—as in our example *a collie*. If definite determiners head their own XPs, then it seems reasonable to assume that indefinite determiners do too. Zamparelli proposes that a predicate nominal is a different type of determiner phrase, suited for introducing indefinite determiners, as illustrated in Figure 10.8.

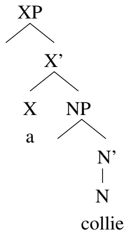


Figure 10.8: Zamparelli's analysis of a predicate nominal construction

Another place where the idea of non-referential nominals finds application is in an account of **existential sentences** such as the following:

(10.19) There is *a dog* in my garden.

There is a well-known similarity between predicate nominals and the nominals introduced in existential sentences. One interesting point of similarity is that there are restrictions on the types of determiner which can be used in both these nominal constructions, and—to a first approximation—the subsets of determiners are the same. The indefinite determiner *a* can be used in both existentials and predicate nominals, as already shown. So can the plural indefinite determiner *some*, as shown below:

(10.20) These dogs are *some beauties.*

(10.21) There are *some dogs* in my garden.

So can 'cardinal' determiners indicating particular numbers:

(10.22) These dogs are *two real beauties.*

(10.23) There are *ten dogs* in my garden.

On the other hand, definite determiners cannot be used:

(10.24) ?This dog is *the collie.*

(10.25) ?There is *the dog* in my garden.

(Example 10.24 is not ill-formed, but it cannot be interpreted as predicating a property of the dog. Instead it functions to state that two separate referential DPs happen to *co*-refer. Example 10.25 is also not ill-formed, but it cannot be read as an existential introducing a new referent: the whole point about *the* is that its referent is *presupposed.*) The class of determiners which can appear in existential sentences are termed **weak**; those that cannot are termed **strong** (Milsark, 1974). Semantically, weak determiners are assumed not to contribute referents (see e.g. Diesing, 1992). Zamparelli supports this semantic analysis with a syntactic proposal: that strong determiners are introduced by a DP, while weak ones are introduced by a different type of determiner phrase, specialised for weak determiners.

What semantic role does a weak determiner phrase play? One thing to note is that while weak determiners do not contribute referents, they do contain information about grammatical number. Predicate nominals can be singular or plural, and existential sentences can introduce singular or plural nouns, as the examples above show. In English, this information is signalled as an inflection on the noun, as well as by the weak determiner. Zamparelli proposes that the weak determiner phrase is the projection where this information is introduced. Accordingly, we can name this projection 'NumP'.[5] To explain how the information is signalled in an inflection on the noun, we can once again assume some form of head movement—for instance, that the noun is generated together with its number inflection, and then raises covertly to the head of NumP to check this inflection.

Of course, 'strong' DPs contain number information too. The English defninite determiner *the* happens not to distinguish between singular and plural, but the noun it introduces can be either singular or plural. And in other languages, different definite determiners are required for singular and plural—e.g. French *le* ('the-singular') and *les* ('the-plural'). Zamparelli proposes that a strong DP takes a NumP as a complement, which in turn takes an NP complement, as shown in Figure 10.9. N contributes a type (in Figure 10.9, 'dog'). Num contributes number information (in the figure, *singular*). And D contributes a referent (in the figure, a presupposed one).

An additional proposal in Zamparelli's model is prompted by the observation that in some contexts, a nominal introduced by an indefinite determiner *can* function referentially. Some examples are given below:

---

[5] Again, the terminology is Ritter's, not Zamparelli's, but the idea is the same.

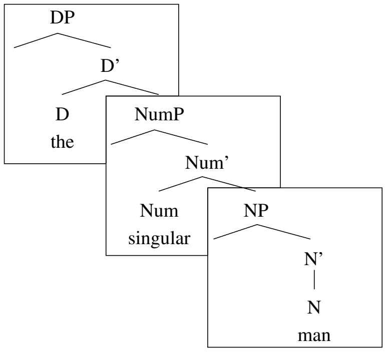


Figure 10.9: Zamparelli's analysis of a strong DP

(10.26) *A dog* came in.

(10.27) Mary gave me *a dog* for Christmas.

We have to assume the presence of a DP in these referential nominals, as well as of NumP and NP. Zamparelli proposes that the indefinite determiner generated at the head of NumP must raise to the head of DP in order to sanction this otherwise empty position. This type of movement is motivated by the same considerations as Longobardi's movement of Italian nouns to empty determiner positions. The movement Zamparelli proposes is of a determiner, not of a noun. But it is nonetheless a form of head movement, in which a head in a nominal construction raises to a higher determiner position.

10.4.2 Ritter's argument for NumP

While Zamparelli's argument for the NumP projection combines syntax and semantics, Ritter's argument is mostly syntactic: it is a development of her analysis of possessive nominals in Hebrew. As well as the construct state possessive form discussed in Section 10.2.2 (see Example 10.28), there is a **free genitive** possessive form, which is illustrated in Example 10.29.

	ahavat	dan	et	acmo
(10.28)	love	Dan	ACC	himself
	'Dan's love of himself'			

	ha-ahava	sel	dan	et	acmo
(10.29)	the-love	of	Dan	ACC	himself
	'Dan's love of himself'				

The free genitive differs from the construct state in two ways: firstly, there is an overt definite determiner *ha-* (realised as a clitic on the head noun *ahava*); secondly, there is an overt genitive case marker *sel* introducing the 'subject-like' element (just as the overt case-marker *et* introduces the object-like element). Note that the free genitive preserves the characteristic 'noun subject object' order of the construct state. Ritter argues that the subject and object elements in the free genitive appear at the specifier and complement of an NP, just as in the construct state. (And for similar reasons: the object can be a reflexive pronoun whose antecedent is the subject.) The head noun must originate at the head of this NP to assign subject and object roles, but must again raise to a higher head position. However, in the free genitive it cannot raise to the head of DP, because there is already an explicit (definite) determiner. Ritter therefore postulates an intermediate XP in between DP and NP, which is the landing site for movement. The obligatory appearance of the overt genitive case-marker *sel* is support for this analysis. (Recall that in Ritter's account of the construct state, the empty D is what assigns genitive case to the subject. If D is occupied by a determiner, an overt case-marker is needed to identify the subject as the possessor.)

As in Zamparelli's account, the precise role of the newly posited intermediate XP is established through a somewhat separate argument. Ritter also argues that the intermediate XP conveys grammatical number information. The argument here is quite simple. Hebrew Ns can have suffixes signalling number; for instance *-im* signals plural for masculine nouns. The suffix on a noun must either be provided in the lexicon, as part of its lexical entry, or it must be contributed independently—which in the GB/Minimalist paradigm requires a separate functional head. If the suffix combines in an idiosyncratic way with the N, then it is likely to originate as part of its lexical entry. But if it combines productively with the N, it is likely to be contributed compositionally, by a separate element. Ritter's main concern is to demonstrate that number suffixes do indeed contribute productively with N (while another type of suffix indicating gender does not). In a way, once we accept the GB/Minimalist idea that semantically productive inflections are contributed by their own functional heads, the existence of a NumP projection follows quite trivially. Seen in this light, the main contribution of Ritter and Zamparelli's arguments is to bolster the case for inflections originating at separate functional heads.

10.4.3 Grammatical number features

As an addendum to the above discussion of the NumP projection, I should note that there is an important difference between grammatical number *features* and actual numbers, of the kind which are denoted by 'numerical' determiners (e.g. *one*, *five*, *ten*). Number features are signalled syntactically, by a closed-class set of words or inflections. The head of NumP is assumed to contribute a particular number *feature*. Syntactically, the determiner which sits at this head must agree with this feature. But it could conceivably supply additional cardinality information of its own. For instance, the determiner *ten* is syntactically plural, and therefore can appear at a NumP head introducing the plural feature, which 'checks' plural morphology on its complement noun. The fact that the determiner also denotes a specific cardinality is a matter of its lexical semantics, rather than its syntactic type.

How much number information do grammatical number features provide? In English, number features distinguish between singular and plural; i.e. between one object and groups of more than one object. In several other languages a slightly finer-grained distinction is made, between singular, *dual* and plural. A dual number feature is found in Arabic, and also several Polynesian languages (e.g. Tongan, Māori); see Veselinova (2008) for more data. There are a few languages which appear to have a number feature signalling 'exactly three', but these are quite rare. The basic number distinctions found in the syntax of natural languages are between one, two and 'more than two'.

10.5 A model of the perception of individual objects and groups

Is there anything in the sensorimotor system which corresponds to the NumP projection? I will argue that there is. In this section I outline a more detailed model of attention to and categorisation of objects which provides the background for this argument. The refined model covers the perception of groups of objects as well as of individual objects. The model was originally presented in Walles *et al.* (2008) and Walles *et al.* (2010); see these papers for details about the implementation.

10.5.1 Group classification in the inferotemporal cortex

If objects in the visual field are of different types, then they must probably be attended to and categorised one at a time. But in the special case where several objects are reasonably close together, and are all of the same type—for instance, a group of trees in a field or a group of cups on a table—observers can attend to a group of objects *as a group*, and classify them *as a group*. I will refer to such groups as **homogeneous** groups. In language, it is easy to refer to a homogeneous group of objects: this is done with a plural common noun. The stem of the noun (e.g. *tree*, *cup*) identifies the common type which the objects share. The plural number morphology (*-s* in English) indicates that we are dealing with a group rather than with an individual. Of course, not all plural common nouns denote visually perceivable groups of objects. But I will begin by considering how homogeneous groups of concrete, visible objects are perceived.

It has been known for a long time that homogeneous groups of objects can be perceived collectively. This fact was a particular concern for the Gestalt psychologists of the early 20th century. These psychologists investigated what properties lead a group of objects to be identified and as a single 'figure'; they identified 'proximity' and 'similarity' as two important properties (see e.g. Koffka, 1935/2001). In modern perceptual psychology, parallel perception of groups falls partly under the topic of 'visual search', and also partly under 'texture perception'. In visual search, it is known that locating a target object in a field of homogeneous distractors is only marginally dependent on the number of distractors (Duncan and Humphreys, 1989). On the assumption that the distractors must be categorised in order to establish that they are not the target, this result shows that a homogeneous group of objects can be categorised in parallel, regardless of how many objects it contains. Our ability to recognise and segregate visual textures is also in part an ability to perceive homogeneous groups. A texture is a visual region in which each point has the same characteristic visual structure (see Landy, 2001 for a good introduction). We can often recognise substances by their textures: for instance, sand or grass have characteristic textures. In some cases, the repeated elements of visual structure can also be categorised as objects (e.g. a grain of sand or a blade of grass), so we expect the boundary between categorising textures and categorising homogeneous groups of objects to be somewhat blurred. This does seem to be the case in the visual system: while there appear to be encodings of textures in relatively early visual areas (e.g. V4, Arcizet *et al.*, 2008), there is also evidence that IT represents natural textures as well as discrete object categories (Köteles *et al.*, 2008). But for the moment I will focus on groups whose homogeneous elements are also clearly objects in their own right.

How are homogeneous groups of objects categorised in parallel? While normal object categorisation happens in inferotemporal cortex (IT), there is some interesting evidence that the output of IT is relatively insensitive to the *number* of objects being categorised.[6] Nieder and Miller (2004) presented monkeys with homogeneous groups of 1-5 objects. The groups varied in cardinality, but also in configuration and retinal position; monkeys performed a task in which the cardinality of groups was the important property. During this task, neurons were recorded from several areas, including anterior IT. While many IT neurons were found to be sensitive to particular configurations of groups, very few were found which encoded cardinality independently of configuration. In another interesting study, Zoccolan *et al.* (2005) found that the response of an IT cell to multiple objects of different types approximated the average of its response to the individual objects by themselves. If IT cells which have this 'averaging' response function also abstract over spatial location, they will be insensitive to the difference between one object of a given type and multiple objects of that same type: in other words, they will be 'blind' to the cardinality of a homogeneous group of objects. In summary, our ability to categorise homogeneous groups of objects in parallel is likely due in large part to spatial abstraction mechanisms within IT. Moreover, the populations of IT cells which can categorise homogeneous groups of objects are probably blind to the cardinality of such groups, responding the same way to one object of a given type and to multiple objects of that type.

Walles *et al.* (2008) built a model of the IT object classifier in which cardinality-invariance emerged as a side-effect of spatial abstraction mechanisms. The model was a convolutional neural network, similar to others developed previously (e.g. Le Cun *et al.*, 1992; Mozer and Sitton, 1996), in which layers of units combining visual features alternate with layers abstracting over local regions of space. After training with regular back-propagation, the output units were able to recognise simple shapes, regardless of their spatial location. The interesting thing was that output units were also relatively insensitive to the cardinality of a homogeneous group of shapes. For instance, an output unit which responded well to a single X responded equally well to a group of several Xs, as shown in Figure 10.10. On the other hand, the trained system did not respond well to heterogeneous groups, which tended to be classified as 'unknown' (see again Figure 10.10). These results

---

[6] Maybe mention the work of Alvarez (2011) on averaging groups here?

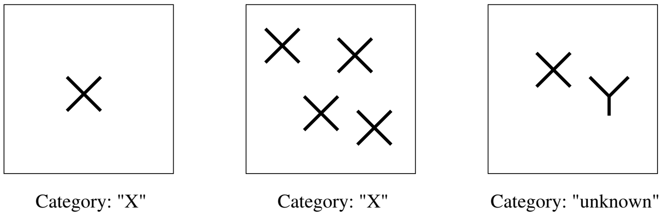


Figure 10.10: Response of Walles *et al.*'s classifier to a single X, a group of Xs and a heterogeneous group

suggest that cardinality blindness might be a natural property of any classifier which is insensitive to the spatial location of objects.

Ecologically, group classification seems to make sense: it is much more efficient to classify a homogeneous group of objects in parallel than individually. Often, the type of object present is behaviourally more important than the cardinality; for instance, a single wolf and a pack of wolves probably require selection of the same basic 'fight-or-flight' response. However, if we revise the role of IT to include the classification of homogeneous groups, we must also revise our model of the attentional visual pathways which run parallel to IT, to specify how they support IT in this role. Attention helps select the input to the IT classifier, and it also helps interpret its output. On the input side, if IT classifies homogeneous groups, then attention must be able to *select* homogeneous groups as salient regions for categorisation by IT, as well as single objects. On the output side, if IT is relatively insensitive to the cardinality of homogeneous groups, there must be some other visual system which delivers cardinality information separately—since observers can undoubtedly tell the difference between a single X and a group of Xs. In the remainder of this section I will begin by reviewing what is known about neural representations of number. Interestingly, representations of number appear to be computed within the attentional system. I will then introduce a revised model of the attentional system presented in Walles *et al.* (2010) which can select groups as well as individual objects, and which can distinguish between singular objects and plural groups.

10.5.2 Neural representations of number

There has been a great deal of research into cognitive representations of number, and the neural mechanisms which implement these. The consensus is that there are two distinct mechanisms for representing number (see Feigenson *et al.*, 2004 for a review).[7] One mechanism computes an approximate representation of the cardinality of a set of objects, called a **numerosity** representation. The defining feature of a numerosity representation is that it supports ratio-based distinctions between cardinalities, allowing two sets to be distin-

---

[7] Naturally I am only thinking about 'pre-mathematical' representations of number, which do not need to be taught. Mature mathematical representations of numbers, such as the concept of an 'integer', supervene on these in ways I will not consider.

guished if the ratio between their cardinalities is sufficiently large. The ratio threshold for a noticeable difference changes with age: for instance, 6-month-old infants can distinguish between 8 and 16 objects, but not between 8 and 12 objects (Xu and Spelke, 2000), while 10-month-olds can distinguish this latter ratio, and adults can distinguish even smaller ratios, on the order of 7:8 (Barth *et al.*, 2003). Importantly, the numerosity system does not seem to work on very small sets, with less than around 4 numbers. An altogether different mechanism appears to compute the cardinality of these small sets. In this mechanism, the numbers 1, 2, 3 and 4 are represented as independent entities, which can all be distinguished from each other. This system is often referred to as the **parallel individuation** system. There are several experiments which demonstrate dissociations between these number systems. For instance, Starkey and Cooper (1980) found that 2-4-month-old infants can distinguish between 2 and 3 objects, but not between 4 and 6 objects, even though the ratio of these larger numbers is the same. Other experiments suggest that the two number mechanisms have separate interfaces to action and/or decision systems, which develop at different rates. For instance, Feigenson *et al.* (2002) presented 10-month-olds with choices between sets of crackers of different cardinalities. When cardinalities were in the range 1–3 (e.g. 1 vs 2 or 2 vs 3), infants reliably chose the larger quantity. But when comparisons involved larger numbers (e.g. 2 vs 4 or 3 vs 6), infants chose at random, even though the ratios were identical—and even though they can already *distinguish* these ratios at this age.

Evidence for these two basic number systems has also been found in non-human primates. For instance, Feigenson *et al.*'s study is actually a replication of a study of monkeys with similar results (Hauser *et al.*, 2000). The neural basis for the numerosity system in monkeys is becoming quite well understood. I have already mentioned Nieder *et al.*'s (2004) study of cardinality representations in monkeys (see Section 10.5.1). Recall that in this study, monkeys were presented with small groups of objects of different sizes and 'global' configurations, and performed a cardinality-matching task. There were not many cells in IT which responded to a group of a particular size irrespective of its configuration (i.e. global shape). However, in other brain areas, high proportions of cells of this kind were found. They were particularly common in the intraparietal sulcus—an area very much involved in the attentional system, as discussed in Section 2.4. These cells appeared to respond to groups of specific cardinalities, regardless of their global or local form.[8] A similar result has been found in humans using an fMRI habituation paradigm (Piazza *et al.*, 2004). The neural basis for the parallel individuation system for small numbers is less well understood. It probably also has a basis in visual attention. There have been several attempts to relate it to a system which tracks small numbers of moving objects in parallel, maintaining separate representations of these objects (see e.g. Trick and Pylyshyn, 1994). I will defer discussion of this system until Chapter 11.

Of course, what we are mainly interested in is a system which represents *grammatical* number, which as mentioned in Section 10.4.3 draws a distinction between singular (indi-

---

[8] In fact, the highest proportion of cardinality-sensitive cells was in prefrontal cortex—a fact which we will return to in Section **??**.

viduals), dual (pairs of individuals) and plural (groups of two or more individuals). What is the mechanism which classifies a visual stimulus into these categories? There are some interesting recent results which bear on this question. For one thing, humans are not the only animals capable of distinguishing between singular and plural: Barner *et al.* (2008) have conducted experiments which suggest that monkeys are capable of making a discrete distinction between these two categories. This suggests that the grammatical singular-plural distinction may have its origin in a prelinguistic system. It has also been found that infants are able to distinguish between singular and plural sets in manual search tasks before they learn to interpret singular and plural morphological features in language (Li *et al.*, 2009). This again suggests that the singular-plural distinction is not just a feature of language, but part of a more basic prelinguistic representational system. Finally, Carreiras *et al.* recently conducted a study investigating fMRI responses to noun phrases containing violations of grammatical number (e.g. *a dogs*).[9] They found that these stimuli caused particular activation in the right intraparietal sulcus, the spatial/attentional area which is involved in computing numerosity according to the review above. This study suggests that the prelinguistic number representations which support the linguistic singular-plural distinction may be representations in the parietal system.

Exactly what these number representations are is still quite an open question. Numerosity by itself does not seem appropriate, since it cannot represent the concept ‘singular’. Parallel individuation is also insufficient by itself, because it cannot represent arbitrarily large numbers. In the next two sections I will introduce a revised model of visual attention, which contains a new proposal about the attentional mechanisms responsible for the singular-plural distinction. In Section 10.5.3 I describe a visual attentional mechanism which can select groups as well as objects. In Section 10.5.4 I describe how this mechanism can interact with the group classification mechanism described in Section 10.5.1 to deliver discrete representations of singular and plural.

10.5.3 A revised model of the attentional system, with selection of homogeneous groups

The idea that attention can select homogeneous regions as well as single objects is well established as an empirical fact: as already noted, similarity is one of the Gestalt principles which group visual stimuli together into a single ‘figure’. In this section I will summarise a simple model of the role of similarity in selecting salient regions, developed by Walles *et al.* (2010).

Walles *et al.*’s model is a modification of the saliency map of Itti and Koch (2000). As outlined in Section 2.4, Itti and Koch’s saliency map is computed from a set of simple visual feature maps, which model the information extracted from the retina in visual areas V1-V4. A salient region is defined as a region of high ‘local contrast’ in one or more

---

[9] The actual NP stimuli were in Spanish, and the subjects were mature Spanish speakers. The selected nouns had irregular plural morphology, to ensure that the activation was due to ‘processing’ of number information rather than to unusual surface forms.

feature maps: in other words, as a region whose features differ sharply from the features of neighbouring regions. On the face of it, it seems difficult to extend this definition of salience to include regions with 'high homogeneity', because homogeneity is exactly the opposite of local contrast. (If two neighbouring regions have high local contrast, then surely they have low homogeneity, and vice versa!) However, Walles *et al.* suggest that the apparent incompatibilities between homogeneity and local contrast as cues to salience can be reconciled if they are assumed to apply at different spatial scales. Recall that feature maps parse the retina at a range of spatial scales: there are fine-grained maps, identifying high-frequency features, and coarser-grained maps, identifying lower-frequency features. In Walles *et al.*'s model, a salient region is one which shows high homogeneity at a relatively fine spatial scale (indicating that its texture elements are homogeneous), but contrasts with its surroundings at a coarser spatial scale. The argument turns on the fact that the 'neighbourhood' of a region picks out different things at different spatial scales. If you pick your scales correctly, then at a fine-grained scale, the neighbours of a point in a given region are (mostly) in that same region, while at a coarser scale, they will mostly be outside the region.

The independent contributions of homogeneity and local contrast to salience in Walles *et al.*'s model are shown schematically in Figure 10.11. The region in Figure 10.11(a) is 'optimally' salient, because it contrasts from its background at a low spatial frequency, but is also homogeneous at a higher frequency. The regions in Figures 10.11(b) and (c) are somewhat less salient: the region in (b) has low-frequency contrast, but is heterogeneous at the higher frequency, while that in (c) is homogeneous at the higher frequency, but lacks low-frequency contrast.

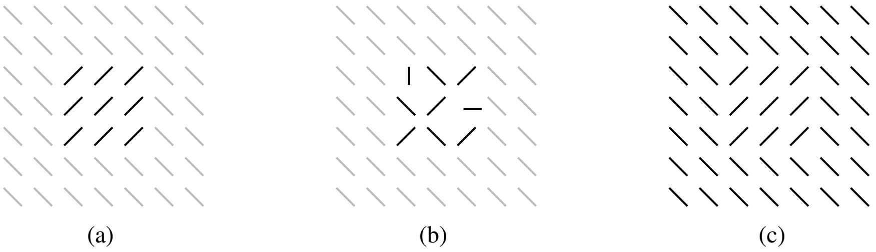


Figure 10.11: Homogenity and local contrast as independent cues to salience

10.5.4 An attentional model of the singular/plural distinction

Walles *et al.* (2010) combine their model of attention with a cardinality-blind classifier. The attentional model is involved in selecting salient regions as input to the classifier: it can deliver regions containing single objects, but also regions containing several objects, if they meet the homogeneity / local contrast criteria just described. However, the attentional system also supplements the output of the classifier, by providing some simple cardinality information, namely a distinction between singular and plural objects. In this section, I will describe how the attentional system provides this information.

One important point about the classifier is that it can operate at two different spatial frequencies, just as the saliency map does. Its inputs are the same maps of visual features as the saliency map uses, so it can work with both high-frequency features and low-frequency features. During training, the classifier is presented with each shape at two different sizes. It learns to use high-frequency feature maps to classify small shapes, and lower-frequency maps to classify larger ones, and thereby achieves 'scale-invariance' as well as location invariance, as illustrated in Figure 10.12. This treatment of scale-invariance is quite standard in models of classification (see e.g. Riesenhuber and Poggio, 1999).

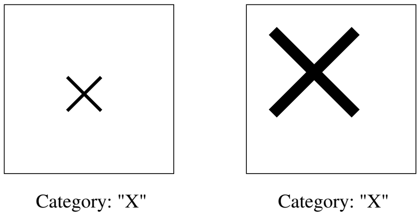


Figure 10.12: Scale-invariance of Walles *et al.*'s classifier

The new idea in Walles *et al.*'s model is that the attentional system selects not only which regions of space the classifier operates on, but also which spatial scale it should use as input. Selecting a 'classification scale' consists in enabling a set of feature maps of a particular frequency: the only features which enter the classifier are those from the selected maps, at the selected retinal location.

Initially, the selected classification scale is a function of the size of the selected region. Walles *et al.* assume that for a region of a given size, there is a **default classification scale**, which is the scale most suited for recognising the primitive components of the shape of an object occupying that region. The idea is illustrated in Figure 10.13. It is obvious that very large visual features are not able to identify interesting elements of the form of the object, and the same is true for very small features. The default classification scale need not pick out exactly one spatial frequency, but it should pick out some fairly narrow range of frequencies to use as input to the classifier.[10]

Figure 10.13: The idea of a 'default classification scale'

---

[10] It is possible that the dimensions of the salient region may select different frequencies for different

Walles *et al.*'s attentional system identifies the size of the currently selected region, and enables a set of feature maps of the appropriate scale, as shown in Figure 10.14. In this system, after the classifier has been presented with the selected region at the default classification scale, there are two possible attentional operations. One is, as usual, to choose a new spatial location to attend to (via inhibition-of-return in the saliency map; see Section 2.4.5). The other is to choose a *new classification scale* at which to analyse the current location, *without shifting location.* The idea here is that there is a special attentional operation which permits a reanalysis of the currently attended location, using primitive features of a different spatial frequency. The new classification scale must obviously be finer than the default one, because primitive features at the coarser scale will be too big to exist in interesting combinations within the selected location. The question is: what can the classifier find at a classification scale finer than the default scale? We argue that at this scale it is optimally configured for *group* classification. The finer classification scale is too fine to detect the component features of a single object occupying the selected region. But if the region happens to contain a group of objects, it will be well suited for detecting the component features of *individual objects in this group.* And if all the objects are of the same type, the classifier will be able to identify this type through the group classification mechanism discussed in Section 10.5.1. In Walles *et al.*'s system, if a region contains a group of objects, it is quite likely that they are of the same type, because the attentional mechanism is predisposed to treat homogeneous groups as salient, as just discussed in

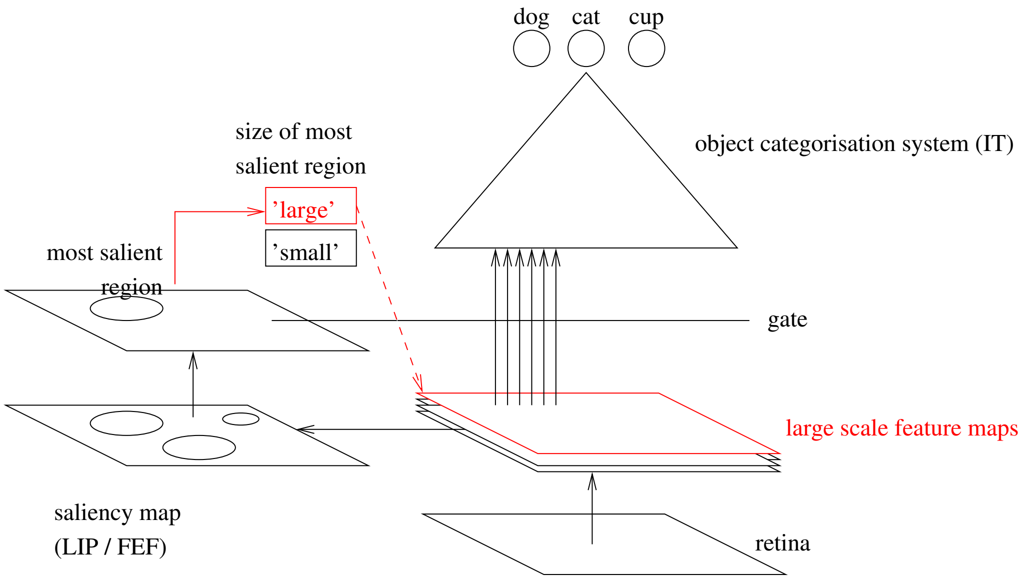


Figure 10.14: Spatial attention and scale attention. The size of the most salient region determines a default classification scale

---

orientations of visual feature. For instance, to classify a tall thin shape, we may want to select horizontally oriented and vertically oriented features of different grain sizes.

Section 10.5.3.[11] The appropriateness of a finer-than-default classification scale for group classification is illustrated in Figure 10.15. The scale which was too fine for identifying

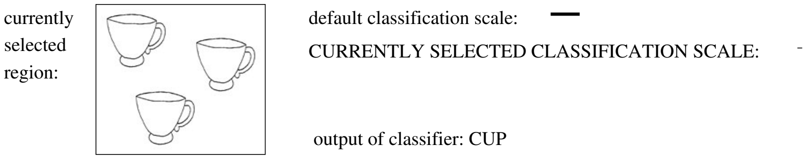


Figure 10.15: Group classification using a classification scale finer than the default scale

the primitive features of an *individual* cup occupying the region (c.f. Figure 10.13) is well-suited for classifying the primitive features of the individual cups in a group of cups occupying the region.

Note that in Walles *et al.*'s system, the configuration of the attentional mechanism when the classifier returns a positive result provides information about the cardinality of the classified object. If the classifier is operating at the default classification scale, there is a single object of the type it has identified. If it is operating at a finer scale, there is a group of objects of this type. Because adjacent spatial frequency channels are around an octave apart (see e.g. Wilson and Bergen, 1979), it is not normally possible to obtain precise information about the cardinality of a classified group. It may be that the immediately higher classification scale is optimal for classifying exactly two objects, but aside from this, the only information that can be read from the setting of the classification scale is a distinction between one object and many objects. In summary, Walles *et al.*'s attentional model naturally divides categorised visual stimuli into singular, dual and plural.

Walles *et al.*'s model of classification is supported by several lines of evidence. The idea of a default classification scale is supported by a well-known attentional phenomenon called **global dominance** (Navon, 1977). The phenomenon relates to stimuli such as that shown in Figure 10.16, in which a large shape of one type is formed from a homogeneous group of smaller shapes of a different type. (In this case, the stimulus is a T made up of Xs.)

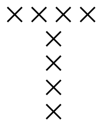


Figure 10.16: Example of a 'Navon' stimulus: a 'T' made up of 'X's

In the classic experiment, subjects watch a series of such stimuli, pressing a button when

---

[11] In fact, attention can also select heterogeneous groups, if they have sufficiently high local contrast. If this happens, the classifier will not be able to identify a type at the finer-grained classification scale. In this scenario, Walles *et al.*'s system attends to each object in the group in turn, using a hierarchical notion of saliency maps which will be explained in Section 10.8.3.1.

a particular shape appears at a particular spatial frequency. Subjects looking for a high-frequency shape tend to be distracted by the target shape appearing at a low frequency, but subjects looking for a low-frequency shape tend to be able to ignore the target shape appearing at a high frequency. Navon argued that observers have a bias for identifying the 'global' shape over the 'local' shapes from which it is formed. Models of global dominance tend to make reference to the notion of spatial frequency channels, arguing that lower frequency channels are given precedence during processing (see e.g. Hughes *et al.*, 1996). Walles *et al.*'s notion of a default classification scale is very much in line with these models. In fact, Walles *et al.*'s system naturally manifests global dominance. Assume that the system has been trained to recognise both Ts and Xs, at a range of different sizes. If it is shown the stimulus in Figure 10.16, it will first select the default classification scale, and the classifier will return 'T'; then it will select a finer classification scale, and the classifier will return 'X'. Walles *et al.*'s system also reproduces some well-known findings about the effects of stimulus similarity in visual search. Duncan and Humphreys (1989) found that search targets were found faster when the distractors were similar to one another. Walles *et al.*'s system reproduces this effect by grouping similar distractors, and classifying them as groups when searching for the target.

It is interesting to note that in Walles *et al.*'s model, information about the cardinality of an attended group (i.e. whether it is singular or plural) is computed outside inferotemporal cortex, in an area involved with visual attention. As discussed in Section 10.5.2, there is evidence that grammatical number agreement features activate pre-linguistic representations of number in intraparietal cortex, but the representations of 'numerosity' which are known to exist in this area do not quite provide the categorical singular-dual-plural information which linguistic agreement features convey. Walles *et al.*'s model suggests how the attentional operations in the intraparietal area could deliver a more discrete singular-dual-plural distinction to the linguistic interface.

10.5.5 Additional sequential structure in Walles *et al.*'s perceptual model

Note that Walles *et al.*'s extended model imposes some additional constraints on the sequential order of operations in the perceptual system. In the basic model of object perception summarised in Section 10.3, there was one ordering constraint: a location needed to be be selected in the saliency map before the classifier could deliver a category (Ordering Constraint 1). In Walles *et al.*'s model, we must add that there is a second attentional operation which must occur after a location has been selected, and before a category can be delivered: the operation of selecting a classification scale. As discussed in Section 10.5.4, the default classification scale depends on the size of the selected region, so a salient region must be selected before a classification scale can be selected. This constitutes another ordering constraint:

Ordering constraint 16 *A salient region must be selected in the saliency map before a classification scale can be selected.*

Moreover, in the revised perceptual model, the classifier cannot deliver any output until a classification scale has been selected. (If it did, the system as a whole would be unaware of whether it was perceiving a single object or a group of objects of the reported category.) Which is another ordering constraint:

Ordering constraint 17 *A classification scale must be selected before the classifier can deliver a category.*

In summary: in Walles *et al.*'s revised perceptual model, perceiving an object or a group of objects involves (at least) three operations, which must take place in strict sequence: first a salient location must be selected, then a classification scale must be selected, and finally a category can be activated.

10.5.6 A sensorimotor interpretation of the extended DP

In Section 10.4, I outlined an extended model of the syntax of DPs. In Section 10.5 so far, I have given a revised model of object perception, supporting the perception of homogeneous groups as well as of individual objects. We can now consider whether the new perceptual model provides a good basis for interpreting the new syntactic model.

Recall that the main extension in the syntactic model was the addition of a NumP projection in between DP and NP, which introduces the grammatical number feature (e.g. singular or plural) of a DP. Is the new model of object/group perception helpful in supporting a perceptual interpretation of the extended DP structure? I will argue that it is. In fact, there are a number of very neat correspondences between the perceptual and syntactic models.

Firstly, note that the way information is decomposed in the perceptual model continues to mirror the way it is decomposed in the extended account of DP syntax. In the syntactic account, type information and number information are delivered separately, by separate projections: NumP and NP. In the perceptual model, type and number information are also delivered separately, by separate visual systems: the classifier delivers type information but not number information, while the attentional system delivers number information but not type information. In fact, the perceptual notion of cardinality blindness is a very natural way of modelling the semantic contribution that a noun stem makes on its own, without a number inflection. In formal semantics, this contribution is quite complicated to state. For instance in Zamparelli's account, the uninflected noun stem *dog* contributes 'the set of all possible sets of dogs'—i.e. the powerset of dogs, an extremely complex object!—from which the NumP then selects those sets with an appropriate cardinality. The assumption of a cardinality-blind classifier makes the semantic contribution of bare noun stems much easier to state.

Secondly, the number distinctions which the syntactic system is capable of making appear quite similar to those which the attentional system is capable of making. NumP distinguishes singular, dual and plural, but very rarely distinguishes between cardinalities any higher than two. In the model presented in Section 10.5.4, the attentional system uses a notion of relative classification scale to make number distinctions. In Walles *et al.*'s implementation, relative classification scale can distinguish between one object and many objects. I suggested that relative scale can probably identify dual as a distinct category as well. But spatial frequency channels are too far apart to allow many more subtle distinctions.

The two correspondences just noted suggest that we can give a sensorimotor characterisation of the NumP projection. This could be stated as follows:

Principle 12 *The NumP projection in a concrete referential DP describes the selection of a relative classification scale in the perceptual system.*

- *The feature 'singular' at Num indicates choice of the default classification scale.*
- *The feature 'plural' at Num indicates choice of a higher classification scale.*

Note that this sensorimotor interpretation of NumP not only makes sense in its own right: importantly, it also makes the right predictions about the internal syntactic structure of DP, given our general interpretation of a right-branching structure of XPs as a description of a sequence of sensorimotor operations. In our extended syntactic model, DP introduces NumP as a complement, which in turn introduces NP. And in our extended perceptual model, selection of a salient region precedes selection of a classification scale, which in turn precedes activation of a category. If DP describes selection of a salient region and NP describes activation of a category as I proposed in Section 10.3, and if a right-branching structure of XPs describes a sequence of sensorimotor operations, then interpreting NumP as describing the selection of a classification scale perfectly predicts the right-branching structure of a DP, with NumP in between DP and NP. This is a particularly significant correspondence, because the sensorimotor interpretation of right-branching XP structures was originally proposed in the context of clause syntax. The fact that it also appears to apply within DP syntax suggests that it may have some degree of generality.

In summary: there is a very natural sensorimotor interpretation of the extended model of DP structure argued for in Section 10.4, which extends the general sensorimotor interpretation of LF structure which was proposed in Chapter 5. It is illustrated in Figure 10.17. DP describes the selection of a saliency map location. NumP describes the selection of a relative classification scale. And NP describes the activation of a category in the object classification system.

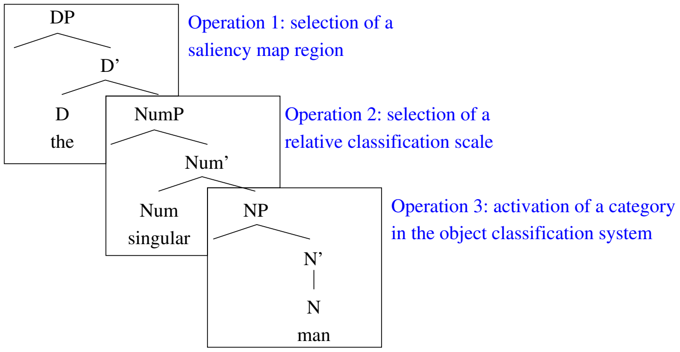


Figure 10.17: A sensorimotor interpretation of the extended DP, including NumP

Of course, our interpretation of DP syntax is far from complete, even for simple referential DPs. For one thing, we need a sensorimotor interpretation of the syntactic operation of head-raising within the DP, which plays an important part in our account of DP syntax. Again, our sensorimotor interpretation of clause syntax in Chapter 5 furnishes us with some predictions. In that chapter, I proposed that head-raising in the clause (from V to Agr to I) was a reflection of the fact that LF describes sensorimotor sequences 'as replayed from working memory', rather than as they occur in real time. If is the case, the presence of head-raising in the DP leads us to make some strong predictions about the existence of a system of working memory for objects (and groups), and about the architecture of this system. For another thing, we need a more thoroughgoing sensorimotor interpretation of 'referents'. As discussed in Section 10.2.3, indefinite determiners 'contribute new referents to the discourse', and definite determiners 'presuppose existing referents'. In the next section, I will review some current models of working memory for objects. In Section 10.7 I will argue that these models do indeed provide the kind of operations needed to give a sensorimotor interpretation of head-raising within the DP—and moreover, that these same models also provide the basis for a sensorimotor interpretation of reference and presupposition.

10.6 A model of working memory for objects and groups

In Chapter 3 I proposed that an observer maintains a working memory representation of the sequence of sensorimotor operations involved in a reach-to-grasp action. The working memory representation can hold a planned action prior to its being executed; it can also hold an action after it has been executed (or observed), when it functions as a buffer used to store the action in long-term memory. In this section, I will discuss our working memory representations of objects and groups. I will suggest that these representations also take the form of prepared sequences: specifically, that an object or group is represented in working memory by the sequence of attentional actions which was involved in establishing it. In Section 10.6.1 I will introduce several different roles which working memory object representations must fulfil. In the subsequent sections, I will consider these in turn, and discuss how they are interrelated.

10.6.1 The function of working memory object representations

It is useful to begin by thinking what purpose could be served by maintaining a representation of an object or a group in working memory. In this section I will outline three functions working memory representations of objects or groups might have.

Firstly, maintaining a working memory representation of an object or group can help an observer *reattend to* this object or group at some point in the future. It is common for an observer to attend to an object, and then reattend to it shortly afterwards (see e.g. Ballard *et al.*, 1997). When we reattend to an object, our perceptual processing is subject to a top-down bias: we expect it to have the same properties as it had before, in some sense (as I will discuss in Section **??**. These expectations constitute a form of working memory for objects. Thinking about a working memory object representation as a device to help re-establish an object gives it a procedural character somewhat like a prepared action. To hold an object in working memory in this sense is to be in a particular 'task set', in the sense introduced in Section 2.6.2, in which we are prepared to execute certain attentional or perceptual operations if given the right cue. Note that location is a particularly important attribute of the working memory representations which facilitate reattention to objects. If an observer's attention happens to be drawn to a location where an object was recently encountered, this should reactivate a representation of that object as a top-down cue to re-establishing it. Conversely, if an object representation is activated top-down as a search cue, this should direct attention to the location where that object was most recently seen.

A working memory representation of an object can also be thought of more declaratively, simply as a means of representing the properties of an object which is not immediately present to the senses. We can evoke a representation of an object which we are not currently perceiving; when we do so, we can be said to be activating a 'working memory' representation of that object. The working memory representation of an object must activate some of the same properties which are activated when the object is before the senses. (Though it need not represent all of these, and it can presumably represent many non-perceptual properties as well.)

Another way of thinking about a working memory object representation is as an intermediary between raw perceptual representations/operations and long-term memory representations. In Section 3.6.4 I proposed that an object is represented in LTM as an atomic neural assembly, and that it is as an atomic assembly that its relationships with other objects is encoded in memory. If objects are established through attentional sequences, something needs to map these sequences onto atomic LTM object representations.

Finally, working memory representations of objects can be thought of as responsible for the fundamental properties of objects in our conceptual system, as entities which persist through time. Again, location is of central importance in this conception of objects. Over short time intervals, an object is *defined* by its location, so that a change in the perceptual properties associated with that location is normally interpreted as a *change in the object*, rather than as the appearance of a new object. Objects can change location, but they normally only do so gradually. The working memory representations which underlie the spatiotemporal continuity of objects in our conceptual system will be discussed in Section 11.2.

10.6.2 Neurophysiological substrates for working-memory object representations

As outlined in Sections 3.2–3.4, there is good evidence that the prefrontal cortex has an important role in maintaining working memory representations of actions. The prefrontal cortex is also the main locus for working memory representations of objects. In this section I will review evidence for the role of PFC in working memory object representations.

It has been known for some time that the PFC evokes representations of of the locations and categories of objects in the visual field, and maintains these representations even after the objects are removed from the visual field if they are relevant to an ongoing task; see e.g. Wilson *et al.* (1993). More recently it has been found that there are PFC neurons which appear to store combinations of category and location information. For instance, Rainer *et al.* (1998) presented monkeys with a preview stimulus consisting of a particular object appearing at a particular retinal location. The object disppeared during a delay period, after which a probe object appeared; the monkey had to give a certain response if the probe object matched the preview object in both category and location. During the delay period, 81% of PFC neurons were sensitive to the object's category or location or both. Of these neurons, 46% encoded a combination of category and location, 49% encoded only the object's location, and the remaining 5% encoded only the object's category. This finding is consistent with fMRI studies in humans showing a partial overlap between the frontal regions involved in object location tasks and those involved in object identity working memory tasks (see e.g. Courtney *et al.*, 1998). In summary, PFC seems to be involved in storing working memory representations of the location and category of recently perceived objects, and of their combination. A very similar conclusion is reached from fMRI studies of human PFC (see Haxby *et al.*, 2000).

It has recently been discovered that some PFC cells are sensitive to the *cardinality* of a visually presented group of objects. As mentioned in Section 10.5.2, Nieder and Miller (2004) found cardinality-sensitive cells in the intraparietal sulcus of monkeys. Earlier, Nieder *et al.* (2002) also found cardinality-sensitive cells in prefrontal cortex—in fact, they were found in larger proportions in PFC (32% of PFC cells studied showed sensitivity to cardinality). When a group of objects was presented, a cardinality-sensitive representation emerged first in the parietal area and then in the PFC, suggesting that parietal cortex computes a sensory representation of cardinality, which is then elaborated in PFC. Nieder and Miller's task required the monkey to compare the cardinality of two groups of objects presented sequentially, interleaved by a delay period. They found that both parietal and

PFC neurons maintain a representation of number during the delay period; however, the memory representation was stronger in PFC, suggesting that PFC is the primary locus of working memory representations of cardinality. The relatively high proportions of PFC cells encoding cardinality (32%) and category and/or location (81%; see Rainer *et al.*) make it plausible that some PFC cells encode cardinality in combination with category and/or location; however, this has not yet been explicitly tested. Neither is it known whether the distinction between singular and plural is represented in terms of relative spatial frequencies, as proposed in Section 10.5.4—although PFC certainly maintains working memory representations of spatial frequencies (see Greenlee *et al.*, 2000).

PFC can also store 'prospective' working memory representations of objects—i.e. representations of objects the agent is looking for, or anticipating (see Rainer *et al.*, 1999). As for actions, working memory for objects can hold forward-looking representations, as well as encodings of the recent past.

10.6.3 WM individuals and their link to LTM individuals

A representation of an object (or group) needs to combine information about location, cardinality and category. It appears that PFC is the place where this information is first brought together when an object is perceived. How does PFC encode cominations of location, cardinality and category? And how are these combinations linked to the LTM representations of individuals which feature in episodic memory? In this section I will argue for a particular model of working memory representations of singular and plural groups, in which they are represented as prepared attentional sequences. This model is a logical extension of the account of object perception given Section 10.5, and of known properties of PFC.

Recall from Section 10.5 that the process of attention to and categorisation of an object has sequential structure. I suggest that the working memory representations of objects developed in PFC encode this sequential structure, rather than representing objects as flat combinations of attributes. This proposal can be motivated in several different ways. For one thing, we have already seen that PFC is well suited for storing observed and planned sequences of sensorimotor operations (c.f Section 3.2)—so there are good reasons to think PFC has the capability of representing objects using a sequential code. Moreover, in order for working memory representations of objects to support reattention to objects (c.f. Section 10.6.1), it is important for them to encode a particular order of attentional operations. For instance, say an observer has recently established a collection of cups on a table, and now wishes to reattend to this collection. The observer must establish the location of the cups before he can establish the appropriate spatial frequency, because the spatial frequency is defined in relation to the size of a selected region in the saliency map (see Section **??**). Similarly, the observer must establish the appropriate spatial frequency before generating a top-down bias towards the object category 'cup'—otherwise the bias will be applied while the observer is categorising the group of cups *as a group*, i.e. establishing what configuration they are in. Finally, and perhaps most importantly, the sequential order of operations associated with a particular object carries information about the nature of the object. This is particularly clear for complex objects. For instance, there is a difference between an X made of Os and an O made of Xs. We need a representation of the properties of an object which makes this difference explicit—a simple combination of the categories X and O will not work. The problem is very analogous to the problem discussed in Section 3.7, of how to represent a transitive action featuring two objects in a way which identifies which is the agent and which is the patient. I propose a similar solution in this case: an object or group is represented in working memory as a planned sequence of sensorimotor operations, which echoes the sequence of representations generated when it was originally established. The planned sequence enables later actions of reattention to the object. I also assume that these planned attentional sequences have a role in creating LTM representations of individuals and groups, as discussed in the next section. I will call the sensorimotor sequence plans associated with objects or groups **working memory individuals** or **WM individuals**. (The term 'individual' is intended to cover both single objects and groups of multiple objects.)

In Section 3.2 it was proposed that planned reach-to-grasp actions are stored using a mixture of competitive queueing and associative chaining. If working memory representations of individuals are held as sequence plans, it is likely that similar mechanisms are involved. Using the graphical notation introduced in Section 3.2.3.3, we could represent a simple individual—say a single dog—as shown in Figure 10.18.

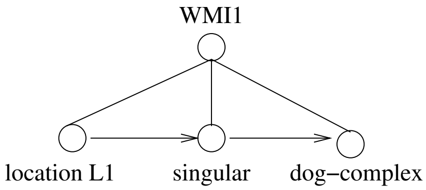


Figure 10.18: A WM individual representing a single dog at location L1

Note that a WM individual encodes a 'property complex' rather than just a single dominant category. In other words, a WM individual carries a rich representation of the perceptual properties of an object as evoked in the property complex layer of the object classification system. This may include perceptual properties which are temporarily hidden, but can be inferred via their association with visible properties. The notion of a property complex is discussed in more detail in Section 12.2.

10.6.4 WM individuals for supporting bottom-up actions of reattention to objects

The model of WM individuals given in Section **??** associates WM individuals with locations: establishment of a location is the first attentional operation stored in a WM individual. However, it is also important to associate locations reciprocally back to WM individuals, in order to support actions of attention driven bottom-up, when a particular location becomes salient. Say an observer's attention is drawn bottom-up to a certain location. If he has recently established an object at this same location, his working memory of the object should exert a top-down bias on his new attentional processes, so that *reattending* to the object is easier than attending to it for the first time.

There is good evidence that reattention is influenced by working memory object representations. A well-known experiment was conducted by Kahneman *et al.* (1992). In this study, observers were shown a preview display containing two boxes, each containing a different letter; see Figure 10.19. During a linking display, the letters disappeared for a short time interval. Then during a test display, a letter appeared in one of the boxes, which the observer had to identify as fast as possible. There were three different conditions. In the 'same-object' condition, the letter to be identified reappeared in the same box it was in in the preview display. In the 'different-object' condition, the letter to be identified was in the other box in the preview display. In the 'no-match' condition, the letter was one which did not feature in the preview display. Observers were fastest at identifying the letter in the same-object display, and slowest at identifying the letter in the no-match display. The different response speeds suggest that observers' object categorisation mechanism can be influenced by top-down expectations created by the preview display. Kahneman *et al.* propose that there are two forms of expectation. One is a general priming effect (see Section 2.2.2), which explains why responses to an unseen letter are slower than to either letter that appeared in the preview display. The other is a mechanism which associates letters with *particular objects*, which explains why responses are faster in the same-object condition than in the different-object condition. This experiment is evidence that reattention to an object invokes a working memory representation of that object, which biases the perceptual mechanisms which re-establish it.

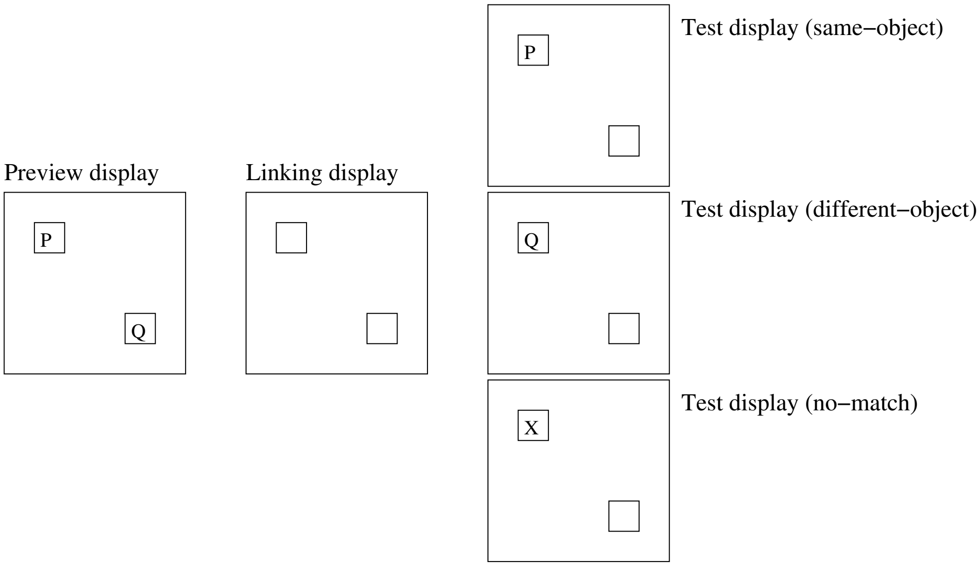


Figure 10.19: Kahneman *et al.*'s (1992) object reviewing experiment (static condition)

The fact that a location can activate a working memory object representation necessitates a revision of the model of WM individuals proposed in Section 10.6.3. Location has a special status for WM individuals. A WM individual can be thought of as a planned attentional sequence whose first operation is to activate a location; however, a location can also activate a WM individual by itself—in which case the first operation in the planned sequence is automatically achieved.

A revised model of WM individuals is shown in Figure 10.20. In this model, locations are

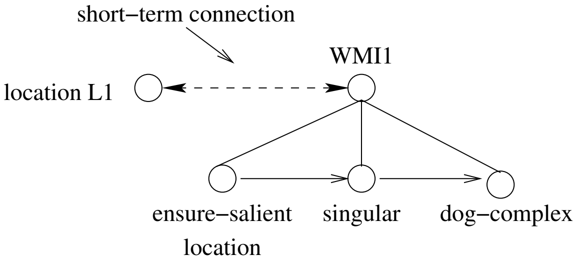


Figure 10.20: A (revised) WM individual representing a single dog at location L1

bidirectionally connected to WM individuals, via short-term connections. (I assume these are established through regular Hebbian learning, when a particular location and sequence plan are simultaneously active.) The first operation in a WM individual's sequence plan is the minimal operation of ensuring that a salient location has been established. If a WM individual is activated bottom-up by a location, this will already be achieved. If it is activated top-down, the associated location will be established through its direct association with the WM individual, and the first operation simply enforces that this operation occurs before the second operation is executed.

The model just shown can account for the object-specific preview effects observed in Kahneman *et al*'s experiment. When a box containing 'P' is seen in the preview display, a WM individual is created representing a cardinality (singular), and a category complex (registering the presence of a box shape and the letter P); this WM individual is associated with a certain location. During the linking display, this association is maintained, even after the letter disappears. In the test display, if a letter reappears in the box, the observer's attention is drawn back to the box's location, which reactivates the WM individual, so that its re-establishment is informed by top-down expectations about cardinality and property complex. If the letter is 'P', these expectations speed letter identification; if not, they slow it down.

There are two issues in the above account which I have skirted over, which will be dealt with in much more detail later in the book. Firstly, I have not discussed the most interesting finding in Kahneman *et al.*'s experiment, or the theoretical proposal which Kahneman *et al.* propose to explain this finding (which is really what the paper is about). The interesting finding is that the object-specific preview advantage is maintained even if boxes *move to a new location* during the linking display. In other words, working memory object representations are not tied to locations—or at least, not exclusively so. Kahneman *et al.* suggest that there is a dynamic form of working memory for objects, which allows an object representation to be associated with a region of space which can be *tracked* if it moves. This working memory medium is termed an **object file**. The concept of an object file is quite influential, and I will be making much use of it in the next chapter; the notion will be introduced in detail in Section 11.2.

Secondly, note that WM individuals are able to represent the properties of objects even if they are not currently visible. to represent the 'P', even though it disappears during the linking period. As already mentioned, I am assuming that a WM individual represents a property complex, which may include properties which are not currently visible (see Section 12.2.3). Presumably the sudden disappearance of the P is interpreted as an 'epistemic change' rather than as an actual event of the P 'leaving' the box.

10.6.5 Summary

10.7 A revised sensorimotor characterisation of DPs

10.7.1 DPs describe replayed attentional sequences

10.7.2 An account of head movement within the DP

10.8 Extensions to the perceptual model to support 'N-of-N' constructions

10.8.1 The syntax of 'N-of-N' constructions

Note: N-of-N constructions are not partitives.

There are three types, illustrated below:

(10.30) On the bridge was *a line of soldiers.*

(10.31) *One of the dogs* barked.

(10.32) I like *this kind of dog.*

10.8.2 A sensorimotor interpretation of the basic N-of-N construction

Basic insight: a line of soldiers is a Navon stimulus. Nothing new needs to be proposed in the perceptual model.

Proposal: *of* says something about changing spatial frequency without changing location. (Still needs to be worked out.)

10.8.3 The *one-of* construction: representing the relationship between a group and its members

These guys are partitives! They're different (or at least, possibly different).

In this section, I will consider a complex DP headed by the determiner *one*, as illustrated below.

(10.33) One of the dogs barked.

This construction is a natural way of reporting a particular perceptual situation, in which we have established a group of dogs by group classification, and we then attend to an individual dog within this group. I will begin in Sections 10.8.3.1 and 10.8.3.3 by introducing a model of hierarchial attentional processing, in which the current salient region can be reparsed as a whole saliency map in its own right, so that individual 'components' of the stimulus identified in this region can be attended to. In Section 10.8.3.4 I will give a model of LTM representations which supports this model of hierarchical attentional processing. Finally in Section 10.8.3.5 I will draw on these models to propose a sensorimotor interpretation of the *one-of* construction.

10.8.3.1 Hierarchical attentional processing

This is part of Walles *et al.*'s (2010) model of attention and group classification.

It is likely that saliency maps can be computed not only for the whole visual field, but also for particular objects or elements within it. For instance, we seem able to perform sequential visual search routines using a canvas far smaller than the whole visual field. A typical visual search experiment asks a subject to look at a screen and find a target amidst a set of distractors. The subject is clearly able to ignore any stimuli which are not on the screen; such objects are not even candidates for being distractors. How do we create a saliency map for a particular portion of our visual field? I propose three mechanisms which permit visual search at different scales.

Note that the model of attention introduced in Section **??** assumes two spatial frequencies. However, there are in fact more spatial frequencies at our disposal: as mentioned in Section **??**, there appear to be around four distinct spatial frequency channels in the human visual system. Let us call these frequencies F1...F4, with F1 being the lowest and F4 being the highest.

Secondly, I suggest that saliency maps are computed at several different spatial frequencies in parallel. Recall that the saliency map function uses two spatial frequencies. Given that there are four spatial frequencies, I will assume that there are three saliency map functions, each of which uses a consecutive pair of spatial frequencies (F1/F2, F2/F3 and F3/F4), and thus that there are three saliency maps, which parse the visual field into salient regions at several different scales.[12] Each of these saliency maps has its own winner-take-all layer, delivering a most salient region. Visual search and object categorisation can

---

[12] Note that since each saliency map function spans two spatial frequencies, the regions established in each map can still be a range of different sizes.

be controlled by any of these saliency maps. Thus I will also assume a function which determines at any time which is the **controlling saliency map**—i.e. the one within which inhibition-of-return operates, and whose most salient region is used to gate input to the categorisation system.

Finally, I propose that the saliency maps are hierarchically organised, as follows: the most salient region in the lowest-frequency saliency map *gates retinal input* to the saliency map function at the next frequency up, so that the higher-frequency saliency map only represents a retinal region corresponding to the most salient stimulus in the lower frequency map. To take an example: say the most salient region in the lowest-frequency map corresponds to a dog—i.e. that the categorisation system, if operating on this region at this spatial frequency, will establish the category 'dog'. In parallel, the next-highest saliency map function will produce a saliency map of all the interesting visual stimuli *within* this dog-shaped retinal region. Thus at any time, we can talk about the **controlling saliency map**, but also a **sub-saliency map** established on the dominant salient region at the next-highest spatial frequency, and also a **super saliency map**, whose most salient region corresponds to the region represented by the current saliency map. (Naturally the sub-saliency map is not defined for the higest frequency map, and the super saliency map is not defined for the lowest frequency map.)

10.8.3.2 Spatial map representations in LTM

Say here how there are map-like representations of environments in LTM. Basic idea: when you evoke a saliency map,

10.8.3.3 Switching spatial frequencies

What are the operations which trigger a change in the controlling saliency map? I suggest that there is a competition *between saliency maps* as well as within maps—in other words, that what is attended to is the most active region in the current controlling map, the sub-map and the super map *combined*.

If the winner is in the current controlling map, we have regular inhibition-of-return. If it is in the sub-map, control is switched to the sub-map. I will refer to this operation as **attentionally entering** the currently established object. Attentionally entering an object creates an attentional context in which we can attend to objects which are 'in', or 'of' that object. For instance, we can establish the handle of the cup as an object in its own right, or a design on the cup's surface. I will talk more about this operation of attentionally entering when spatial representations are considered in Section **??**.

If the most salient region is in the super map, control is switched to the super-map. I will refer to this operation as **attentionally pulling back** to the region which embeds the region currently dominating attention.

Note that the operation of attentionally entering the current object must be distinguished from the operation of 'group classification' discussed in Section 10.5.1. In group classification, we switch the object classification system to a higher spatial frequency, but we do not change the controlling saliency map, and we continue to use the current most salient region in this map to gate input to the classification system. When 'attentionally entering' the current most salient region, we switch to a higher spatial frequency in the object classification system, but *also in the controlling saliency map*, which means we focus on a sub-region of the current most salient region. Thus while group classification requires us to attend collectively to all the elements which are 'in' or 'of' the currently established object, attentionally entering the object allows us to iterate through these elements as individuals. Thus while group classification could establish that a bowl is entirely full of cherries, or entirely made of wood, attentionally entering the bowl allows us to attend to individual parts of the bowl, or individual things that it contains.

10.8.3.4 LTM individuals and LTM groups, and their relationship with WM individuals

[This looks okay: the idea is that we rehearse a WM individual to create LTM representations, but that LTM representations match perceptual stimuli (locations and property complexes) directly. However, when you want to do something with an LTM representation—e.g. when you want to read it out—you're constrained by WM processing constraints on adopting 'working spatial frequencies' to do it serially. When you do this, you re-create a WM individual to represent the LTM item.]

Individuals and groups must be represented in long-term memory, as well as in working memory. A simple model of LTM individuals was introduced in Section 3.6.4—however, we must now extend it to provide an account of how groups of individuals are stored in LTM, and to describe how LTM representations of individuals and groups relate to the WM representations of individuals just introduced.

I will begin by revisiting LTM representations of single objects. I will continue to use the term **LTM individual** to refer to these representations. As described in Section 3.6.4, an LTM individual is an assembly in parahippocampal cortex which is linked to a complex of sensory categories, which we can now associate with a 'property complex'. As described in Section 3.6.4.2, LTM individuals are also associated with locations. Together, these two associations allow a familiar object to be 'recognised': if we see an object with a certain property complex, and there is an existing LTM individual which is associated with a sufficiently similar property complex, and the object's location is close enough to the location most recently associated with this LTM individual, then we can re-activate this LTM individual, and assume we are looking at the same object we saw before. Obviously, an object can change both its location and its properties, so an LTM individual's associations with a property complex and with a location must both be gated by context. (However, some properties obviously change more slowly than others—for instance, the size of a person changes over a timescale of years. Properties of this kind must be associated with LTM individuals via links which are less contingent on context, or perhaps contingent on very coarse-grained representations of context which only change slowly themselves.) Note that this model of an LTM individual should allow a simple single object to be recognised without any serial processing: a representation of the current property complex and currently attended location should be able to act simultaneously to access a stored LTM individual. (If there is no such individual, there must be a mechanism which creates one. I will briefly speculate about how this mechanism works at the end of this section.)

Now consider LTM representations of groups of objects. These must somehow be distinct from LTM representations of single objects. LTM group representations must in fact be considerably more complex than representations of single objects. There must be a way of identifying a common property complex associated with each element in the group (so we can represent 'a group of dogs' in LTM), and there must also be a way of identifying the arrangement or configuration of the group (so we can represent specific configurations such as 'a line of dogs'). There must also be a way of identifying the size of the group, either precisely or approximately. Finally, there must be a way of representing the membership of particular individuals (represented as LTM individuals) within the group. The representation of the group must be able to support later recognition of the group, as with single LTM individuals. However, it must also be possible to recognise individual members of the group if they are subsequently encountered by themselves. For instance, say a cat Tibby has two identical kittens, whose properties allow me to distinguish them from any other kittens, but not from one another. If I subsequently see one of these kittens, I must be able to recognise it as 'one of Tibby's kittens', even though I cannot say which of them it is.

To support these requirements, I suggest that LTM representations of groups involve a separate type of assembly, which I will call an **LTM group environment**, or just **LTM group**. (The use of the term 'environment' prefigures an account of spatial environments given in Chapter **??**. In that chapter, I give an account of spatial environments, and argue that a group of objects is a special kind of spatial environment.) I assume that an LTM group is an assembly in parahippocampal cortex, just like an LTM individual. An LTM group is associated with exactly one LTM individual, and the individual and the group can be associated with distinct property complexes. Moreover, an LTM group is associated with a spatial representation: while LTM individuals are associated with locations, an LTM group is associated with a whole *map* of locations, in the medium in which spatial maps are stored.[13] Different cardinalities or numerosities are represented by maps with different structures. In addition, individual points in the map can be associated with LTM individuals, to represent the membership of particular LTM individuals in the group. An LTM representation of a group is as shown in Figure 10.21. The configuration shown represents 'a row of soldiers'. LTMI1 is an LTM individual, representing the row as a single entity, with its own location. It is linked to the property complex 'row', and to a particular location L1 in a saliency map. LTMG1 is an LTM group, representing the row 'as an environment', or 'as a group'. It is linked to the property complex 'soldier', which is the property complex which is evoked when the objects in the group are categorised via group classification. (Thus any property complex linked to an LTM group will indicate the category to which the *elements* of the group collectively belong.) LTMG1 is also linked to

---

[13] This medium is also implemented in the hippocampal area—it is discussed in its own right in Chapter 13.

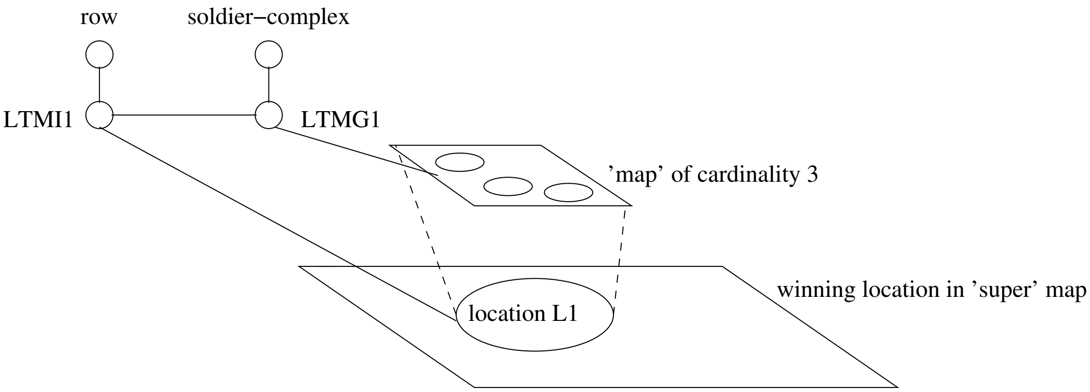


Figure 10.21: An LTM representation of a group: 'a row of (three) soldiers'

a spatial 'map' representation, containing three locations. The map is an LTM structure encoding the saliency map which was activated when the group was established: I suggest that reactivating the LTM group *reactivates this map*, just as it can reactivate a collection of sensory properties. Note that LTMI1 and LTMG1 are reciprocally linked: activating LTMI1 can activate LTMG1 and vice versa. Note also that there is a relationship between the location associated with LTMI1 and the map associated with LTMG1. Recall from Section **??** that the model of saliency maps is hierarchical: the 'current saliency map' is associated with a single location in a 'super-saliency map' at a lower spatial frequency. This same relationship is assumed to hold between the location associated with LTMI1 and the map associated with LTMG1.

The configuration just described is designed to support the LTM representation of groups both in terms of their configuration, the category of their individuals, and in terms of their cardinality or numerosity. Thus in the example shown in Figure 10.21, the configuration is given by the property complex associated with the LTM individual, the category of the component individuals is given by the property complex associated with the LTM group, and the cardinality or numerosity is given by the type of map associated with the LTM group.

What is the relationship between a WM individual and this LTM group representation? I suggest it is somewhat analogous to that between WM and LTM representations of episodes: namely, that the LTM representation is created by a process of *replaying* the WM individual created during perceptual experience, in a special mode where active sensory representations are associated with LTM strucutures. A serial replay operation seems necessary, because there are two property complexes ('row' and 'soldier') which must be associated with different LTM assemblies; if the assemblies and property complexes were all activated simultaneously, many spurious associations would be created (at least, if we assume Hebbian learning). The serial replay operation has the following structure. To begin with, a new LTM individual is created, and associated with the first attentional operation stored in the WM individual: activation of a location. The next operation is the activation of the default spatial frequency, which is a ‘null’ operation in the interface with LTM. The next operation is the activation of a property complex, ‘row’, which in the LTM interface is linked to the new LTM individual. The next operation is the activation of the higher spatial frequency, which in the LTM interface has the effect of creating a new LTM group assembly, and associating this with the new LTM individual. The final operation is the activation of the ‘soldier’ property complex, which in the LTM interface is associated with the new LTM group.

While a sequential replay process is needed in order to *create* an LTM group representation, I assume that the process of recognising a known group when it is re-encountered is instantaneous, just as it is for single individuals. Say we have a group of soldiers represented in LTM, as in Figure 10.21, and some time later we re-establish this group. At the point of establishment, our perceptual state will involve an active ‘soldier’ property complex, and a saliency map with a particular cardinality or numerosity. I assume that this state is able to reactivate the LTM group assembly, because it is linked to both these perceptual representations. Note that recognition happens through the ‘soldier’ property complex rather than through the ‘row’ property complex. When establishment is complete, the active property complex is ‘soldier’; ‘row’ is only active transitorily during the establishment of the row of soldiers. This is appropriate, because we must be able to recognise the group of soldiers even if they have changed their configuration and are no longer in a row.

Note that the ‘soldier’ property complex is also activated if the observer sees just one member of the group of soldiers. Establishment of a single member of the group is thus able to activate the whole LTM group. In this case, however, there is a discrepancy between the spatial representations associated with the perceived object and with the activated LTM group: the group is associated with a complete saliency map, while the perceived object is associated with a single location. The full configuration activated in this case is as shown in Figure 10.22. The elements ‘directly’ representing the single soldier in LTM are shown

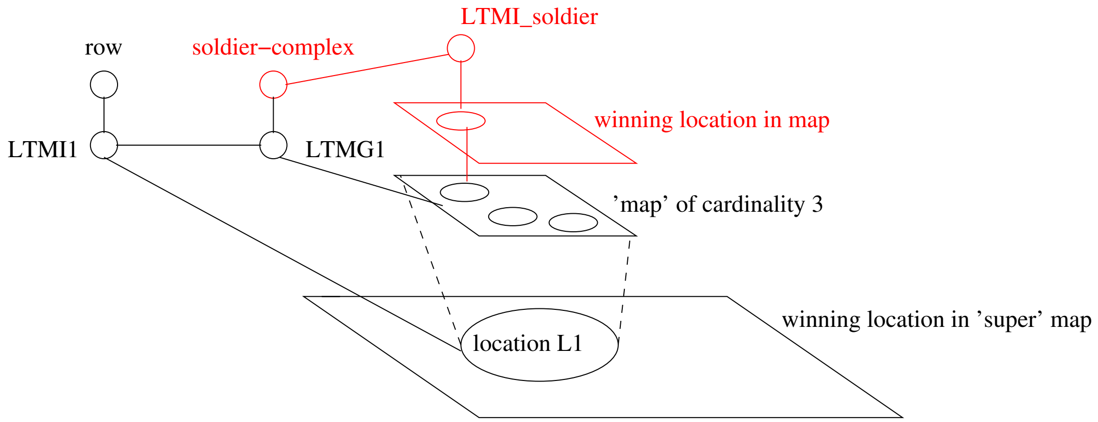


Figure 10.22: The configuration which supports recognition of ‘one of the soldiers’

in red. The key relationship here is between the saliency map associated with the group and the single location representing the individual soldier. In Section **??** I will argue that this relationship is denoted by the preposition *of*. For the moment, the main point is that WM individuals are used to create representations of groups in LTM, by a process of replay similar to that used to create episode representations in LTM.

Is replay of a WM individual also needed to create an LTM representation of a single object—i.e. of an LTM individual? I will assume that it is. Thus if an established individual object is not recognised, a process is triggered whereby the WM individual is rehearsed, leading to the creation of a new LTM individual. To begin with, a new LTM individual is created. This is linked to the first operation stored in the WM individual: establishment of a location. The second operation, choice of the default spatial frequency for categorisation, has a null effect in the LTM interface. The final operation, activation of a property complex, associates this complex with the newly created LTM individual.

10.8.3.5 A sensorimotor interpretation of the *one-of* construction

Again, *of* says something about the relationship between analyses at two (adjacent) spatial frequencies. We get *the dogs* by picking a salient region which encompasses the whole group of dogs. We get the individual dog by moving to the higher spatial frequency. Obviously when we do this, *we don't change category*: we already know what the category is. So while in *A line of dogs* we keep the location and change the frequency and category, in *One of the dogs*, we move to a sub-location, change the frequency, but keep the category.

10.8.4 The *kind-of* construction

Idea about KIP: *of* (occupying the head of KIP) denotes a special operation which can be executed to allow referring expressions to refer to internal cognitive categories / predicates rather than to things in the world.

There's a special process we need to do to refer to a kind.

- First we execute the 'of' operation, to establish the domain of kinds as possible referents rather than the domain of objects in the world.

- Then we have to evoke a regular (nameable) category (e.g. book)

- Then we have to express a regular SDP, which (within the context which is set up) will be understood as referring to a type. It must do so unambiguously—this can be done by pointing to an object (by which we will now understand the type of that object) or by generating an anaphoric reference (which in this special context will pick out a recently-evoked *property*, rather than an object), or by constructing a relative clause which presents a property which uniquely picks out the property (e.g. 'The kind [of book] which you can't stop reading').

I want to interpret this within Roberto's framework. I think the KIP probably describes the point where the property is evoked, and the SC complement of KI is probably 'just referential'. The structure of the SC relates to the structure of the process by which the property is explicitly identified (i.e. is identified in a way which is communicable).

Question: what is there in common between the *of* in this construction and in the previous two? My idea was that *of* says something about switching spatial frequencies while preserving something else. I need to generalise this to get a good meaning for *of* in this case. Here's my idea: when you move from attending to a group to attending to an individual within that group, you *also* get a shift from a type to a subtype. (Because not all individuals in the group are the same: they have their own characteristics.) (There's no other perceptual operation which reliably shifts from a type to a subtype like this, so this is the operation which provides the external characterisation for this operation of internal deixis.) So doing this is a little like exploring the space of subtypes within a type. (TBC..)

10.8.5 The *picture-of* construction: intensional contexts in DP

Another interesting type of partitive is *a picture of a dog*. This is different from *a line of soldiers*. Syntactically, the nested noun has its own determiner (which looks to be a full strong DP actually). Semantically, this nested DP is interpreted in an intensional context, like the sentential complement of verbs like *say* and *believe*. The dog doesn't have to actually exist—it's just a representation of a dog.

What does *a picture [of a dog]* have in common with *say [that P]*? In my treatment of *say* verbs (see Caza and Knott, 2012; Knott, 2014), I suggest that a hearer listening to a speaker's utterance has to process it twice—first to identify it as a particular class of physical action ('talking') and secondly as a meaning-bearing stimulus, whose words are mapped to semantic entities. My proposal is that infants have to learn that talk actions are special in that they *can* carry meaning. Infants learn that when hearing a 'talk' action, they should enter a special mode called 'verbal mode', in which semantic representations are activated by learned associations with incoming word forms, rather than through the normal sensorimotor channels. In our model, this is a piece of operant learning: the action of entering verbal mode is learned through a special reinforcement scheme that rewards the agent's ability to correctly predict the word he hears next. If the agent routinely establishes joint attention with observed speakers, 'talk' actions constitute a good cue to establish verbal mode, and we have a model that learns to do this.

I now suggest that something similar happens when an agent perceives a picture. The agent first categorises the picture as a physical object of a certain type (namely painting, drawing, photograph, or just picture), without identifying any of its content. This categorisation is what is done at the 'default classification scale'. But there's something special about objects of this general type which makes it worth re-analysing them as meaning-bearing stimuli. When you look at a dog, it's not worth asking what it *depicts*: it's just a dog, that doesn't depict anything. But when you look at a painting, it is worth asking this question. (Just as, when you perceive a talk action, it's worth asking what the action is *about*, but not when you perceive someone sneezing.)

What does it mean to decide to ask what a picture depicts? My suggestion is that the boundary of the picture has to be re-interpreted as the boundary of an arbitrary scene—

as if the agent had opened their eyes and could only see this scene, and was obliged to interpret it as the actual world. I'll call this mode 'picture interpretation mode'. This mode of perception is not quite like verbal mode: the agent doesn't engage a completely new set of learned associations between perceptual stimuli and semantic concepts. He is still using his regular perceptual mechanisms. But he has in some sense 'suspended disbelief': a stick figure of a dog will be interpreted straightforwardly as a dog, because he's *looking* for meaning. So there is some kind of new cognitive mode in place when he parses a picture for its meaning. In one sense, however, entering picture interpretation mode is like entering verbal mode: it involves disengaging the current sensorimotor context. Whatever we expect to find in a picture need have nothing to do with the situation the actual physical picture is in. Analogously, what we expect a sentence to be about need have nothing to do with the situation in which it's uttered. In both cases, the 'semantic content' of the stimulus can relate to something far removed in space or time from the physical stimlus.

How does an infant learn to routinely re-parse pictures, photos etc for meaning, but not other classes of object? I like the idea that operant (reward-based) learning is again involved. But I don't think the mechanism can involve successful prediction. And the mechanism can't just relate to successful activation of semantic concepts. (The infant can attend to other actual objects in his environment and get plenty of those.) My proposal is that the things represented by pictures happen to be particularly interesting: more interesting than the things in the infant's actual environment. (Diggers, animals, etc—all the stuff you find in children's books.) Assume infants get a reward for every semantic concept they activate, but that different concepts have different rewards, then if they get a higer reward on average when they re-parse a picture for its meaning than when attending to an arbitrary item in the world. And assume they get a punishment if they re-parse something that's *not* a picture. On these assumptions, I think we can explain how after training they routinely parse pictures for meaning, and not other classes of object.

In *picture of a dog*, what does the word *of* denote? It doesn't signal a special intensional context, since it's also used in *line of soldiers*, *box of cherries* etc. I think it just signals that the stimulus at the salient location is going to be re-classified—but doesn't specify whether this re-classification will target the texture elements of the stimulus (as in *line of soldiers*), its physical contents (as in *box of cherries*) or what it represents (as in *picture of a dog*).

10.9 Summary

Section 10.5 presented an extended model of attention to and categorisation of objects. A key feature of this model was that perceptually establishing an object (or a group of homogeneous objects) is an operation with a characteristic sequential structure, just like perceiving (or executing) a reach-to-grasp action. Section 10.6 developed a model of working memory for objects: WM individuals are stored attentional sequences. (...)

Chapter 11

A sensorimotor characterisation of the DP-clause interface

11.1 Introduction

In this chapter I consider the syntactic relationship between DPs and the clauses in which they appear, and look for correlates of this relationship in sensorimotor cognition. As already mentioned at the start of Chapter 10, the relationship between DPs and clauses is not straightforward: it is made complex by the existence of **quantified DPs**, whose semantic scope extends over the whole clause in which they are embedded. In this chapter, I will show that the relationship between the perceptual mechanisms which support 'attention to objects' and those which 'recognise events' is similarly complex—and moreover, that an analysis of this relationship can shed useful light on the relationship between DPs and clauses.

The main point is that attending to an object taking part in an event is necessarily a *dynamic* process, which is extended in time. Objects undergo changes when they participate in events, and the mechanisms which attend to objects must track them through these changes. I begin in Section 11.2 by discussing some experiments exploring the nature of these dynamic mechanisms of attention to objects, introducing the concept of **object files**—the working memory representations which support these mechanisms. In Section 11.3 I formulate an account of the relationship between the sensorimotor and working-memory systems responsible for attending to and representing objects, and their counterparts responsible for apprehending and representing episodes. In Section 11.4 I turn again to syntax, and outline the key syntactic relationships between a DP and its host clause. In Sections 11.5 and 11.6 I draw the syntactic and sensorimotor accounts together. Again, there are some interesting correspondences to be made.

11.2 Object files: a dynamic form of working memory for objects

We have already discussed the processes involved in attending to objects in some detail, in Section 10.6. The basic model developed in that section was that when an observer attends to an object, he forms a working memory representation—a 'WM individual'—to help him reattend to that same object some time in the near future. A WM individual basically associates a location with a particular collection of object properties. One of the key experiments indicating the existence of working memory object representations was that of Kahneman *et al.* (1992), reported in Section 10.6.4. However, Kahneman *et al.*'s experiment is in fact best known for another condition which I have not yet described. I will first discuss the additional condition, and then consider its implications for models of working memory for objects.

11.2.1 Kahneman *et al.*'s experimental evidence

Recall that Kahneman *et al.*'s (1992) experiment featured a preview display, in which certain letters appeared in certain boxes, followed by a linking display, where the letters disappeared, and then a test display, where a single letter appeared in one of the boxes. In the conditions described in Section 10.6.4, the empty boxes remained stationary throughout the linking display. But there was in fact another condition, where the empty boxes *moved to new positions* during the linking period, along smooth trajectories, as shown in Figure 11.1. Subjects were still faster at categorising the reappearing letter in the same-object condition than in the different-object condition. It thus appears that during the

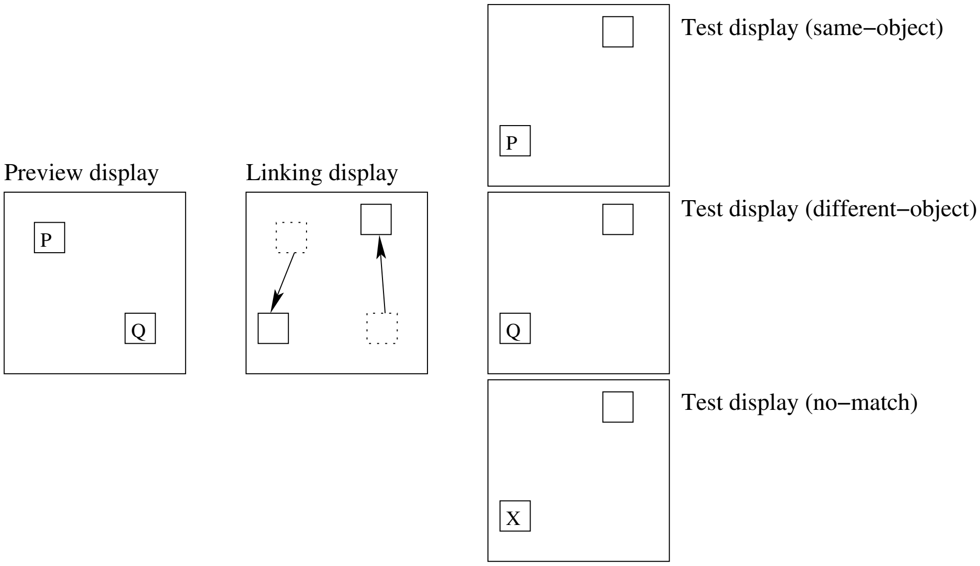


Figure 11.1: Kahneman *et al.*'s (1992) object reviewing experiment (motion condition)

preview display, the observer associates a letter with a representation of the box *as a moveable object*, rather than just as a location in the display.

Kahneman *et al.* introduce the term **object files** to describe these relocatable object representations. When an observer first sees a box in the preview display, he creates a 'file' of information about this box. The file makes reference to the letter contained in the box—say 'P'. It also makes reference to the box's current location. During the linking display, the location associated with the box is updated as it moves; however, the letter P remains associated with the box, even though it is not visible. When a letter reappears in the box, the observer is biased towards re-establishing the letter P, which is why his response is faster in the same-object condition than in the different-object condition.

The fact that preview benefits survive the continuous motion of boxes in the linking display prompts Kahneman *et al.* to propose a more far-reaching model of object files. The key mechanism is the one which updates the location associated with an object file as the visual stimulus it represents moves through the visual field. This tracking mechanism is insensitive to perceptual changes in the stimulus, for instance the appearance and disappearance of letters. Kahneman *et al.* see it as providing the perceptual basis for the cognitive mechanism which maintains an object's identity over time. While a stimulus is being tracked, the observer is attending to a single object—any changes to the location or perceptual characteristics of the stimulus, or to semantic categories reflecting these characteristics, must be interpreted as *changes undergone by the object*, whether these be to its location or to its semantic type.

There are several interesting points to discuss about this idea, which I will consider in the remainder of this section.

11.2.2 Tracking mechanisms and the saliency map

One issue which Kahneman *et al.* do not fully address is the issue of *what is tracked* by the visual tracking mechanism they introduce. It is circular to say that what is tracked is 'an object', because the tracking mechanism is invoked precisely to explain what constitutes an object. Neither is it possible to define what is tracked by specific perceptual features, because the tracking mechanism is defined precisely to be insensitive to changes in such features. Fortunately, the account of visual attention given in Section 10.5 provides scope for a better answer to the question. I suggest that what is tracked is a *region of the saliency map.* An active region in the saliency map indicates that 'something' is present at the corresponding location in the world, but says nothing more about it. We can thus think of an object file as being associated with an active saliency map region, and envisage a tracking mechanism which updates the region associated with an object file when this region changes gradually over time. The perceptual mechanisms which support our concepts of objects and of their continuity through time over changes in location and perceptual characteristics can be clearly expressed at the level of the saliency map.

11.2.3 Multiple object tracking and the implementation of object files

In Kahneman *et al.*'s experiment, there is a same-object preview advantage for both of the objects in the display; it thus appears, *prima facie*, that observers can allocate several object files to different points in the visual field and track their movements simultaneously. Kahneman *et al.*'s experiment is often related to another set of experiments by Pylyshyn and collaborators (see e.g. Pylyshyn and Storm, 1988) which appears to point to the same conclusion. In these experiments, a subject is shown a display of several dots, a subset of which is briefly illuminated. The subject is required to track the illuminated dots (now indistinguishable from the others) while all the dots move around the display. A test dot is then illuminated, and the subject must indicate whether this dot is one of the ones to be tracked. Subjects can reliably track around four or five dots in this paradigm. Pylyshyn has proposed a model of working memory object representations which highlights the importance of tracking processes in individuating objects, and stresses the limited number of such representations which can be maintained at any one time (see e.g. Pylyshyn, 2001). On this model, subjects accomplish the multiple-dot-tracking task by associating each dot to be tracked with an individuating object-file-like representation.

The neural machinery which implements relocatable object representations is not well understood. It is particularly hard to model several independent object tracking mechanisms. Some models make use of the temporal synchrony approach to binding—these models assume that different object files are associated with different phases of a pervasive cyclic neural signal, with their respective locations firing in synchrony (see e.g. Kazanovich and Borisyuk, 2006). However, the mechanism of binding by temporal synchrony is still the subject of much debate. There are also solutions which rely on 'conventional' binding by short-term synaptic connections. Some binding models envisage a collection of 'binding units', each of which is fully connected to all locations and to all object types, which can implement a binding between objects and locations through short-term synaptic weights (see e.g. van der Velde and de Kamps, 2006). The model is illustrated in Figure 11.2. Two object files are shown; each object file is connected to each location. For each object file there is a single strong connection (shown in red) which links the file to a particular location. Again, the neural plausibility of this model is also open to debate. The difficulty in this case is the highly localist nature of binding units. As yet there is no evidence for a small number of distinct neural assemblies with these patterns of connectivity. Moreover, the binding units model only implements static associations between objects and locations.

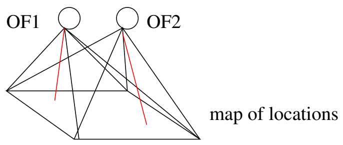


Figure 11.2: A localist model of object files as binding units

It would be hard to extend it to cover associations which are maintained when an object moves to a new location, because this requires the movement not just of activity, but of 'a single strong synaptic connection'. This connection cannot be established using a Hebbian rule, otherwise all active points will become associated with all active object files.

It seems likely that any scheme for implementing object files using conventional binding by synapses must envisage a large amount of neural machinery. I will outline a scheme in which each object file is implemented by *a complete spatial map*, as illustrated in Figure 11.3. Each layer in the figure denotes a map of locations. The leftmost map is the saliency map, which projects in the usual way to a winner-take-all layer, representing the most salient location. The three layers on the right are specialised for object tracking. The most salient location always activates a corresponding region in the **tracking map**, which projects to two further maps, each of which 'implements' a single object file. These maps each project independently, and reciprocally, to the tracking map.

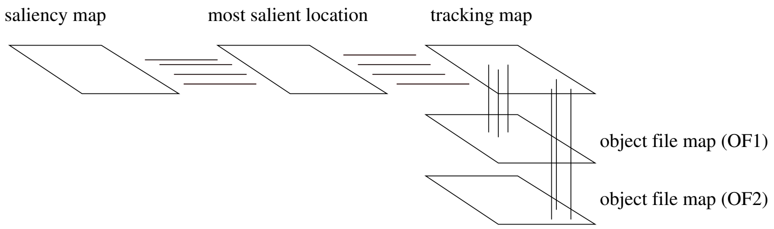


Figure 11.3: Object files as maps

Activity in the tracking map is governed by two principles. Firstly, an active point in the 'most salient location' map always strongly activates the corresponding point in the tracking map. Secondly, activity in the tracking map 'sticks' to the region it was assigned to if this region moves sufficiently incrementally. I will assume a simple model of low-level motion tracking, driven by motion energy. In this model, if motion in a given direction is detected at an active point at a given moment, the active point at the next moment is moved in the direction of this motion. Importantly, a point activated by tracking is less strongly activated than a point activated by focal attention.

Activity in an object file map is also governed by two separate principles. One principle causes an object file to be *initialised*—i.e. generates a new active point in the object file map. Another principle causes an object file to be *maintained* from one moment to the next. To initialise an object file, we must first attend to a location, activating a point in the 'most salient location' map. The active point in this map always strongly activates a corresponding point in the tracking map. We must also 'select' an object file to represent this point. The rule governing initialisation of an object file is as follows: if the activity of a point in the tracking map exceeds a certain threshold, it automatically activates the corresponding point in a selected object file. Say object file 1 is selected to track the currently attended location. The strongly active point in the tracking map will then activate a point in the 'object file 1' map. Maintaining an object file is implemented by the following principle: a point P in an object file map is activated at the current moment if its corresponding point in the tracking map is active, *and* P (or a point neighbouring P) was active at the previous moment. This principle allows an object file to 'stick to' its associated point in the tracking map, even if this point moves (incrementally).

Figure 11.4 shows how object files can be assigned to two regions, which can be tracked independently. The saliency map initially contains two regions, R1 and R2. At time T1, region R1 is attended to, generating a strongly active point in the tracking map. (Strongly active points are shown in red.) I assume that object file OF1 is selected to track this region; therefore the strong activity in the tracking map activates a corresponding strongly active point in OF1. (The connection between these two points will be maintained independently of focal attention, and is also shown in red.) At time T2, the most salient region is inhibited. However, the associated point in the tracking map is still active, though at a lower level. Consequently, the corresponding point in OF1 is still active (again at a lower level). At time T3, the other region in the saliency map, R2, becomes most salient, and OF2 is selected to track this region. The most salient point strongly activates a point in the tracking map, and this point is activated in the OF2 map. (The point associated with R1 is not activated in OF2, because it is below the necessary threshold.) At time T4, focal attention is withdrawn from R2; now R1 is being tracked in OF1 and R2 is being tracked in OF2. Provided R1 and R2 maintain spatiotemporal continuity, and do not get too close to one another, these regions can be tracked separately.

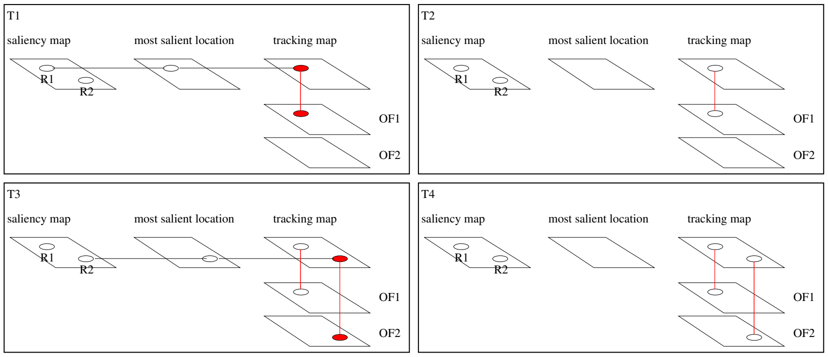


Figure 11.4: Initialising and tracking two separate regions

I do not want to argue too strongly for this model of object files—as already mentioned, the neural basis for object tracking is still quite a mystery. However, there are a few points which speak in its favour. Firstly, it allows a useful dissociation between focal attention and sustained attention. An object can be attended to without being tracked by an object file (if no object file is selected); conversely, once an object is being tracked, focal attention can be allocated to it or removed from it without affecting tracking. Note also that the activity of an object file provides information about both sustained and focal attention: if an object file has an activity level higher than a certain threshold it is tracking an object, and if it has a high activity level, then focal attention is also being allocated to the tracked object. Secondly, the tracking map can do service by itself in an account of multiple-object tracking. It appears that multiple objects can be tracked without being distinguished as tokens (see e.g. Pylyshyn, ??). In the model just given, points can be activated in the tracking map without being assigned to object files; this would result in the observed dissociation between object tracking and object individuation. Finally, there is general evidence that multiple-object tracking involves parietal regions and the motion-sensitive early visual area MT (Culham *et al.*, 1998). The tracking map and object-file maps are likely to be implemented in parietal cortex, and motion tracking is likely to involve MT. However, this evidence is obviously consistent with many possible models of object tracking.

Perhaps the most clunky aspect of the model is that it assumes each object file is a complete spatial map, whose locations are projected one-to-one onto the tracking map. However, this assumption will appear less profligate once a fuller account of cognitive spatial representations has been presented. For the time being, I will assume the current model of object files and multiple object tracking; it will be refined and set in context in Section 11.3.

11.2.4 Object files and the object categorisation function

Finally, it is useful to consider the role that object tracking can have in training the object categorisation function. As noted in Section 2.2, the object categorisation function implemented in IT abstracts over the retinal location, size and (to some extent) orientation of the object being classified. How is this function learned? One very interesting proposal is that the function is constrained to deliver an output representation which only changes slowly over time (Wiskott and Sejnowski, 2002). This simple constraint is sufficient to force the function to deliver object representations which are invariant to the location, size and orientation of a presented object. However, the constraint about slowly changing representations can only be enforced *while a mechanism for tracking a salient retinal region is under way.* The constraint does not apply if the observer abruptly switches attention to a new point in the visual field. Thus the routine which trains the object categorisation function must make reference to the tracking mechanism which which underlies our notion of the spatiotemporal continuity of objects.

11.3 The link between object and episode representations

I have now given a somewhat more elaborate model of the mechanisms involved in allocating attention to objects, and representing them in working memory. In this section, I turn to the question of how these more detailed object mechanisms relate to the mechanisms involved in monitoring 'whole episodes'. I will turn my attention back to the cup-grabbing episode discussed in Chapters 2 and 3. In these chapters, the object representations 'man' and 'cup' were basically placeholders: I did not attempt to investigate the differences between ‘cup’ and ‘cups’, or between ‘a cup’ and ‘the cup’ (let alone those involved in ‘many cups’ or ‘every cup’). In this section, I will suggest how the richer model of object representations developed in Sections 10.5–13.11 of this chapter can be incorporated into a model of episode representations. My aim is to develop a model which is sophisticated enough to serve as the foundation for an account of the syntactic relationship between clauses and DPs.

I will begin by focussing on working memory representations: specifically, on the way WM episode representations interface with working memory representations of objects. I will begin in Section 11.3.1 by making a fairly straightforward proposal about the relationship between WM episode representations and *WM individuals.* In Section 11.3.2 I will discuss the relationship between WM individuals and object files, and propose a more elaborate conception of WM episodes which makes reference to both WM individuals and object files. The remainder of the section will flesh out this proposal, and develop a more detailed proposal about the mechanisms involved in storing an episode in working memory, and in rehearsing the episode to store it in long-term memory or to interface with language.

11.3.1 Extending the model of WM episodes to incorporate WM individuals

Recall from Section 3.2 that a WM episode is a working memory representation of the sequence of sensorimotor signals evoked during experience of an action. It is held in PFC, and is able to be replayed, for instance to create a record of the episode in longer-term hippocampal memory (Section 3.8.1) or to ‘read out’ the episode to a linguistic interface (Chapter 7). I suggested in Section 3.2.3 that WM episodes are stored in PFC using a mixture of competitive queueing representations and associative chaining representations; in the current section I will focus on associative chaining representations. To recap from Section 3.2: the basic structure of a WM episode encoding *The man grabbed a cup* (using an associative chaining representation) is shown in Figure 11.5.[1] The ‘WM episode’ is an assembly of PFC cells which impose certain biases on the pathway from stimuli to

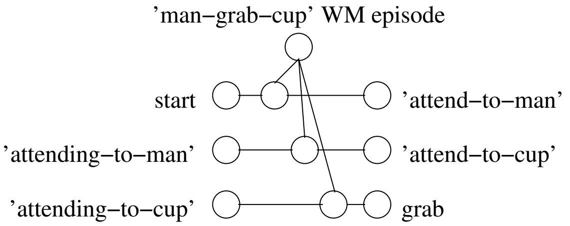


Figure 11.5: The WM episode representing ‘The man grabbed a cup’ (associative chaining notation)

---

[1] I have omitted the recurrent context representations, for clarity.

responses. The initial stimulus is a ‘start’ signal; the WM episode biases the agent to respond to this by activating the operation ‘attend-to-man’, which was the first attentional operation executed when the episode was experienced. The reafferent consequence of this operation is the evocation of the category ‘man’ in IT. The WM episode also contains a bias from this stimulus to another operation, ‘attend-to-cup’; the reafferent consequence of this is the evocation of the IT category ‘cup’. Finally, the WM episode has a bias from this stimulus to the ‘grab’ motor programme. When the WM episode is active, activating the ‘start’ signal will result in a sequence of operations: ‘attend-to-man’, ‘attend-to-cup’, ‘grab’, interleaved with their reafferent consequences.

We now need to rethink the operations ‘attend-to-man’ and ‘attend-to-cup’, as well as their reafferent consequences. As discussed in Section 10.5, the process of attending to and classifying an object has an internal temporal structure of its own, so ‘attend-to-man’ and ‘attend-to-cup’ are each complex operations in their own right. But note that each attentional sequence is itself stored in working memory: as discussed in Section 10.6, a WM individual stores the sequence of attentional actions needed to reattend to a recently attended object (or group). We can thus propose a relatively simple revision to the concept of a WM episode, in which attentional actions are (top-down) activations of WM individuals, as shown in Figure 11.6. In this representation, ‘WMI-man1’ is intended to denote a particular WM individual. When a WM individual is activated, it normally results in an attentional sequence: attention to a particular location, then the establishment of a particular cardinality, then the establishment of a particular property complex in IT. I assume that this sequence can be internally rehearsed, just like a WM episode can be, since it is implemented via similar mechanisms. In addition, a WM individual is linked to a particular LTM individual. In summary, by making reference to WM individuals, a WM episode accesses a much richer representation of objects, permitting the rehearsal of attentional sequences, including representations of cardinality as well as location and type, and interfacing with LTM representations of individuals.

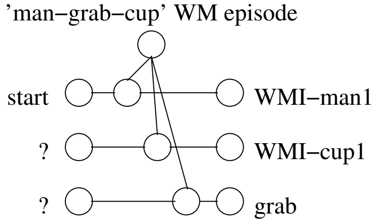


Figure 11.6: A revised WM episode representation, featuring references to WM individuals

A WM episode can now be thought of as a sequence plan containing an element of hierarchy. The first element of the planned sequence is the activation of another planned sequence. When this subsequence is completed, the second element of the planned sequence is executed, which is also the activation of a planned sequence.

11.3.1.1 The new WM episode representation in an account of verb raising

At this point it is useful to recall the 'sensorimotor' account of verb raising given in Section **??**. The phenomenon to be explained was why the inflections of a verb agreeing with the subject and object appear on the verb, rather than in local relationships with the subject and object—and why the inflected verb can raise out of the verb phrase to higher positions. My proposal was that the subject and object agreement inflections denote *planned* actions of attention to the agent and the patient, and that the verb denotes a *planned* motor action. When a sensorimotor sequence is *replayed*, all three planned operations are tonically active at each point during replay, and can therefore be associated with phonological representations at each stage. If they are read out early, then the verb appears out of sequence ('prospectively'); if they are read out late, then the inflections appear out of sequence ('retrospectively').

Note that agreement inflections only carry particular kinds of information about a verb's subject (or object). The most common information carried concerns person (first, second or third), number (singular, dual, plural) and a range of other attributes (gender, animacy, edibility and so on). The account of agreement in Section **??** was based on a simple model of attention to objects, and did not go into details about the information conveyed by agreement inflections. But now that we are using WM individuals to represent planned actions of attention, this account can be considerably expanded. Note that a WM individual holds a tonically active representation of a direction of attention, the establishment of a cardinality, and the evocation of a property complex. It is quite plausible that directions of attention to onesself, to a hearer, or to a third party are encoded differently. It is plausible that singular, dual and 'multiple' cardinalities are encoded differently. And it is plausible that the interface between planned property complex activations and phonological representations only allows certain general object classes to be expressed. In other words, the model of WM individuals developed in this chapter, and the proposal that WM individuals feature in WM episodes, fit very well with the sensorimotor interpretation of verb raising developed in Section **??**. The revised model of WM episodes forms the basis for a very good account of verb raising and verb inflections.

11.3.1.2 Reafferent consequences in WM episodes

Note that the reafferent sensory consequence of an action of attention is left unspecified in Figure 11.6. How should the end of a successful attentional sequence be signalled? It is not possible to refer to a WM individual which has been reactivated bottom-up, because at any time there may be a large number of these, representing the different objects which the observer has recently established. Moreover, if a WM episode consists of biases to arbitrary pathways between pairs of WM individuals, this requires a combinatorial explosion of pathways. Later in this section I will propose a different solution, which draws on the other variety of object working memory introduced in Section 10.6: object files. But first I must address a question which was deferred in Section 10.6: how the notion of object files relates to the notion of WM individuals.

11.3.2 The role of WM object representations in action monitoring: overview

As summarised in Section 10.6.5, I introduced two quite separate conceptions of working memory object representations. A **WM individual** is a static representation, which serves to link the sequence of perceptual operations required to establish an object at a single point in time to a location, and to a LTM individual. An **object file** is a dynamic representation of a currently attended object over a continuous period of time, which is associated with a special low-level tracking mechanism, and which supports our conception of the spatiotemporal continuity of objects, and of events involving their displacement or change. In the remainder of Section 11.3, I will outline a model of how these two types of object representation are related. The key idea in the model is to reinterpret 'object files' as representations that play a role in the sensorimotor processes involved in the execution and perception of *actions*—whether these be reach-to-grasp actions or locomotion actions—and more generally in the experiencing of any episode.

I will begin in the current section by giving a high-level account of how object representations (both in WM and in LTM) are involved in the process of experiencing an action (whether as the agent or as an external observer). As just reiterated in Section 11.3.1, this process is assumed to have a sequential structure. In order to describe the role played by WM and LTM object representations, it is useful to divide the process into three phases.

In the **initial phase**, the objects which will be involved in the action are established one at a time, via focal attention. The first object established is the agent; the second is the target object. During this phase, each object is assigned a WM individual, which stores a static representation of the object's location, and of the object's properties at the current time. This WM individual is linked to an LTM individual—either an existing one (if the object is recognised) or a new one (if it is not). These links are contingent on the observer's representation of 'the current temporal context'. In addition, an *object file* is associated with each object. While the WM individual encodes a static memory of the object's location, cardinality and category at a particular time, the object file maintains a dynamic representation of these properties. When each object is focally attended, its associated object file will be strongly activated (see Section 11.2.3 for the mechanism which implements this), and the LTM individual representing the object it is tracking will also be activated. At this point, a temporary association is formed between the object file and the LTM individual, which lasts for the duration of the action.

In the **action monitoring** phase, the object files track the objects involved in the action as these undergo changes, and the action is categorised. As a side-effect of categorisation, focal attention is returned to the agent. (Recall that the model of object files given in Section 11.2.3 allows for an object to be tracked both with and without focal attention—see especially the discussion of Figure 11.4.)

The final phase is the **consequent phase**, when the action is completed. At this point, focal attention is reallocated to the patient—again, while both agent and patient are still being tracked. The consequent phase involves two operations, which are coordinated in time, and will be discussed at length in Section 11.3.5. One is the operation of replaying a working memory representation of the action to episodic memory, which is itself coordinated with the operation of updating the observer's representation of 'the current temporal context'. The second is an operation creating a *new WM individual* for each object file, representing the object's new properties and location in the new context. The new WM individual might be the same as the one associated with the file in the initial phase; however, it need not be. The new WM individual is also linked to an LTM individual (in the *updated* temporal context). The LTM individual is constrained to be the one associated with the object file in the initial phase, regardless of the object's current properties or location. This constraint implements the axiomatic continuity of objects undergoing actions (see Section 11.2).

The above account can be thought of as an elaboration of the accounts of events as sensorimotor sequences developed in Chapters 2 and 3 and in the current chapter. The new account provides a new *temporal* characterisation of the sensorimotor and attentional states in these sequences. Some of the states involved in representing an action (those established during the initial and consequent phases) are associated with individual points in time, while others (those involved in the monitoring phase) are associated with continuous periods of time. The states are linked in the following way. To begin with, the static attentional states created during the initial phase provide the representations necessary to transition to a dynamic attentional state. Then, when the action has been monitored to completion, the dynamic attentional states created during this process in turn provide the representations needed to transition to another static attentional state.

This temporal characterisation of the sequence of states evoked during action monitoring permits a clear account of how 'static' WM individuals relate to 'dynamic' object-file representations, answering the question posed at the start of this section. During the initial phase, static WM individuals serve to *initialise* the object files which will be used to monitor the action. During the consequent phase, the object file representations serve to create new static individuals to represent the changed objects and link them to appropriate individuals. In summary: by incorporating WM individuals and object files into the existing model of how actions are experienced, we are able to clearly state the relationship between them.

11.3.3 Object files and the concepts of 'agent' and 'patient'

Note that by giving object files a role in the process of action monitoring, the above account adds considerable substance to the notion of object files. When introduced in Section 11.2, object files were primarily motivated by a special form of object-tracking experiment. Their ability to represent objects undergoing change was discussed, but there was no detailed theoretical proposal about their role in perceptual processing. The account just outlined suggests a specific proposal: the role of object files is to *represent the participants in actions* as dynamic entities. According to this proposal, object files are associated with specific participant roles, such as 'agent' and 'target object' (for a reach-to-grasp action), or 'locomotor' and 'target location' (for a locomotion action).

Associating object files with participants in actions allows an account of object files to be grounded in detailed models of action monitoring. We have already given detailed models of the monitoring of reach-to-grasp actions (see Chapters 2–3) and of locomotion actions (see Section 13.11). This association sheds useful light on many questions about object files. For instance, the question of 'how many separate object files there are' can be approached by considering how many separate participants must be tracked when monitoring different types of action. (Pylyshyn 2001 in fact grounds his conception of FINSTs by asking a similar question about event monitoring.) In each of the action types we have considered, the answer is that two object files are required: one to track the agent (or locomotor), and one to track the target object (or location). Accounts of how other actions are monitored may require the postulation of more object files. The point is that the question 'how many object files are there?' can be answered with reference to models of action monitoring, rather than by an arbitrary number deterimed through experiment.

Another advantage of associating object files with participants in actions concerns the implementation of the tracking mechanism which allows an object file to remain associated with a referent over changes to its location and intrinsic properties. Recall from Section 11.2.3 that implementing object files requires the postulation of a separate spatial map for each individual object file. Our models of action monitoring already require separate spatial representations for agent and target object or location: agent must be represented in an environment-centred coordinate system (Section 2.8.3, Section 13.5.3.2), while a target object must be represented in a body-centred coordinate system (Section 13.2.2.2, Section 13.7.1) and a target location must be represented in a specialised map of 'goal locations' (Section 13.11.2). If object files are given specific roles in action monitoring, the postulation of separate spatial maps for each object file is much less unparsimonious than it originally seemed in Section 11.2.3.

A final advantage of using object files to model the concepts of 'agent' and 'patient' concerns the way the object file mechanism encodes a combination of focal and sustained attention. I have argued that monitoring an action involves a sequence of directions of attention, in which focal actions of attention to agent and patient occupy particular serial positions. But I have also argued that agent and patient must be separately and continuously tracked *throughout* the course of an action, in order to represent their spatiotemporal continuity during the action. The object file mechanism provides a way to reconcile these two apparently inconsistent conceptions of agent and patient. As discussed in Section 11.2.3, the activity of an object file carries information about both sustained and focal attention: a certain level of activity indicates that an object is being tracked, while transitory bursts of activity above this level signal in addition that it is being focally attended. These bursts provide opportunities to *associate* particular object files with transitory representations activated at particular points during action monitoring, even while several object files are simultaneously tracking separate objects. To illustrate, recall the important idea that the agent of the cup-grabbing action is reattended to when the 'grab' action is established, and that the patient is reattended to when the action is monitored to completion. Object files can represent these focal actions of reattention, even while agent and patient are continuously tracked. Figure 11.7 shows the time-course of agent and patient object file activation during the experience of the cup-grabbing episode. Note

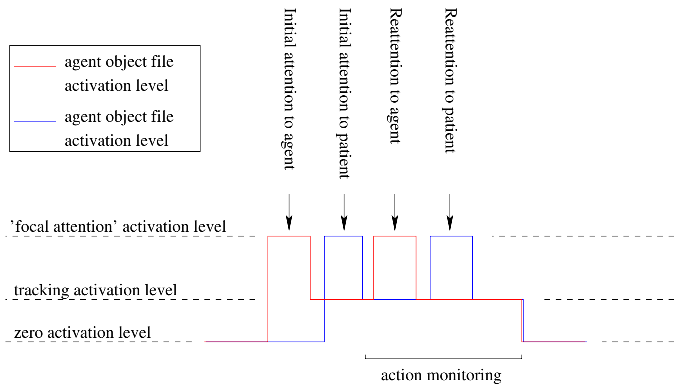


Figure 11.7: Activation levels of agent and patient object file during experience of the cup-grabbing episode

that there are two distinct levels of object file activation; the lower level indicates that an object file has been assigned, and is tracking a salient region; the higher level indicates that it is being attended to, or reattended to. To begin with, the agent is focally attended to, and assigned an object file which then tracks it throughout the remainder of the episode. At the moment it is initialised, it has a high level of activation. (It is at this point that it is associated with a LTM individual.) Next, the patient is focally attended to and assigned a different object file in a similar manner, while the agent object file is sustained at a lower level of activation. Next, the 'grab' action is recognised, and as a reafferent consequence, the agent is reattended to; at this point the agent object file again has a burst of high activity, allowing additional representations to be associated with the agent via regular Hebbian associations. Finally, when the action is complete, the patient is reattended to, and the patient object file is reactivated in its turn, to allow a haptic representation to be associated with the entity it is tracking.

In summary, the object file mechanism allows us to finesse the dual status of agent and patient as sequentially ordered attentional operations and as continuously tracked entities. It is a powerful mechanism for encoding some of the fundamental properties of participants in actions.

11.3.4 References to object files in WM episodes

The proposal that object files are involved in tracking specific participants in actions opens up several other useful new ways of representing actions (and episodes in general) in working memory. If each participant in an action is associated with an object file, these object files provide methods for individuating participants in WM episodes *without reference to their intrinsic properties*. The whole point of an object file is that it is assigned to an individual on the basis of its location alone, and remains associated with this individual as it moves or changes. At the same time, if there are distinct, and specialised, object files associated with the agent of an action and its target object (or target location), then object files are able to identify the participants in an action in terms of *what role they play* in the action. According to the model being proposed, the first object to be attended is associated with a particular object file—i.e. is tracked using a particular spatial map (centred on the environment), and the second object to be attended is associated with a separate object file—i.e. tracked using a separate spatial map (either a body-centred map or a map of goal locations). References to object files—i.e. to specific spatial maps—thus permit references to the participants in an action, in a way which identifies the *roles* of these participants, and which is sustained throughout the course of action monitoring, while abstracting away from their identities.

There are two ramifications of this idea for the represention of WM episodes, which I will enumerate below.

11.3.4.1 Planned actions and reafferent consequences in WM episodes

My main proposal is that references to object files are included both in the representation of the attentional actions involved in a WM episode and in the representation of the reafferent consequences of these actions. When the agent is first attended to, this action can be represented in two ways: first as the activation of a WM individual (as proposed in the previous section), and second as the activation of a particular object file (namely the agent). Moreover, the reafferent consequence of having attended to the agent can be represented as a state in which the agent object file is *active*. The action of attention to the patient can likewise be represented both as the activation of a WM individual, and as the operation of activating the patient object file, and its reafferent consequence can be represented as a state in which the patient object file is active. According to this model, a WM episode for the cup-grabbing episode will be as shown in Figure 11.8. Note that WM episodes now make reference to both WM individuals and to object files—i.e. to both forms of working memory for objects.

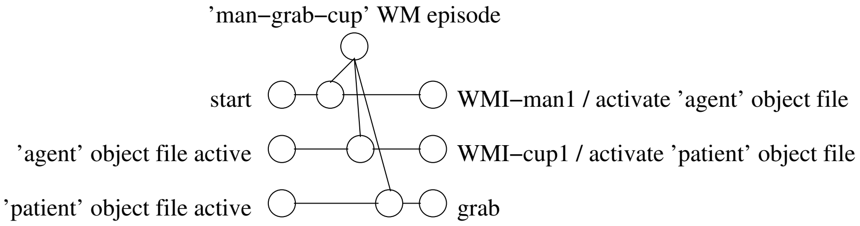


Figure 11.8: A revised WM episode representation, featuring references to object files

Recall that when we extended our model of WM episodes to include reference to WM individuals (see Figure 11.6), we left open the question of how to characterise the reafferent sensory consequence of attending to an object. References to active object files provide a useful way of specifying these consequences. In fact, the WM episode representations which result are very economical in terms of the number of pathway units which are needed. Rather than requiring a pathway unit for each possible pair of WM individual representations, the number of pathway units is linear in the number of WM individuals, and proportional to the (very small) number of participant roles. It is thus quite an efficient way of representing WM episodes.

Note that there is an important difference between the *operation* of activating an object file and the *state* of an object file being active. A WM episode schedules *operations*: the first two operations are actions of attention to two objects, and the third operation is activation of a motor programme. Each action of attention is represented dually, as the activation of a WM individual and as the activation of an object file. Activating an object file results in a state in which it is active. But recall that during experience of an episode, attention can also be allocated to objects as a side-effect of other operations: the operation of categorising the grasp action reallocates attention to the agent, and the completion of the grasp action reallocates attention to the target. Thus while the *operations* of activating each object file each occur only once during experience, each object file enters an active *state* at two separate points.

It is important for each object file to become active twice during replay, to encode the actions of reattention to agent and patient which occur during experience. Since these operations are not scheduled, I think we must assume a special mechanism which activates the agent object file whenever an action category is active, and which activates the patient object file in the state we transition to after activating an action category. This would imply that working memory has an inbuilt propensity to associate actions with an agent object file, and the end state of an action with the patient object file.[2]

11.3.4.2 References to variables in WM episodes

The idea of referring to object files in WM episodes can be taken one step further. It is sometimes important to represent an episode in WM in a way which abstracts away from the details of one or more of its participants—i.e. which specifies participants as 'variables' which can match arbitrary individuals. For instance, a query to LTM might take the form of a WM episode with one participant represented as a variable: *Who grabbed the cup?* abstracts over the identity of the agent, while *What did the man grab?* abstracts over the identity of the target. There are several other linguistic constructions which feature episode representations containing variables, notably quantified sentences (e.g. *Every man grabbed a cup*). I will discuss these constructions in detail in Sections **??** and **??**. For the moment, I will suggest that references to object files provide a means for abstracting over WM episode representations in the appropriate ways. For instance, *X grabbed the cup* can be represented by a WM episode whose first action is simply the activation of the 'agent' object file, with no associated WM individual, as in Figure 11.9(a). It is also possible to abstract over the contents of both participants of an action, to leave a bare specification of the 'argument structure' of the action, as in Figure 11.9(b).

---

[2] I need to introduce this idea much earlier, when the notions of reattention to agent and patient are first discussed.

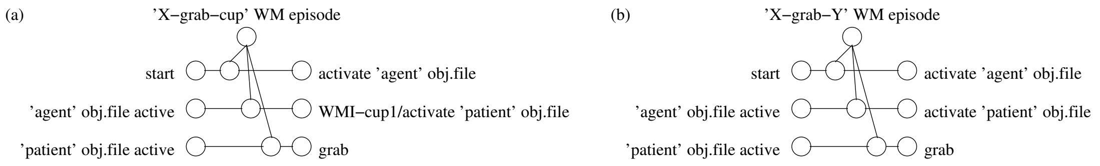


Figure 11.9: (a) WM episode representing 'X grabbed the cup'. (b) WM episode representing 'X grabbed Y'.

11.3.5 The role of object files during replay of a WM episode

One of the key functions of a WM episode is to buffer the sequence of sensorimotor operations involved in experiencing an event or state so that it can subsequently be replayed. Replaying a WM episode allows the associated sequence to be stored in hippocampal episodic memory (Section 3.8.1), or associated with explicit linguistic signals (Chapter 7). In the last few sections, the notion of a WM episode has been considerably revised to include reference to rich working memory object representations. These revisions require us to extend the account of what happens when a WM episode is 'replayed'. The extended account is outlined in this section.

In Section 11.3.5.1 I will introduce the idea that an object file holds a 'memory' of events that occur during experience of an episode. This memory is quite circumscribed, since it relates to a single attentional operation, but nonetheless it contains information which spans the length of the whole episode. In Section 11.3.5.2 I will consider the replay of WM episodes to a linguistic medium. In Section 11.3.5.3 I will consider the replay of WM episodes to episodic memory.

11.3.5.1 Object files and their associated WM individuals

An important idea in the account given in Section 11.3.2 is that each object file involved in monitoring an action is associated with two WM individuals: one represents the static properties of the tracked object in the initial phase; the other represents its static properties in the consequent phase. Introducing some terminology, we can say that by the time an episode has been monitored to completion, an object file has links both to an **initial WM individual** (describing the tracked object as it was in the initial state) and to a **new WM individual** (describing it as it now is, in the consequent state). In this section I will discuss these links in more detail.

The link from an object file to its initial WM individual is quite straightforward to describe. When an object is attended to, a WM individual is created, and an object file is strongly activated: the link between the object file to the individual can simply be thought of as a short-term Hebbian association between an object file and a WM individual. However, the link from an object file to its 'new' WM individual is of quite a different sort. Recall from Section 11.3.2 that when an observer has finished monitoring an episode, the object files which track its participants are involved in 'creating' new WM individuals, representing the new properties of the tracked objects, as these may have changed during the monitoring process. What is involved in this process of creating a new WM individual? I suggest that to enable the creation of a new WM individual, an object file maintains representations of the (possibly changing) cardinality and category complex associated with the tracked object as well as its location. I assume that at the end of an episode, a new WM individual is 'read off' these representations.

Note that this way of creating WM individuals is quite different from the way they are created during a 'regular' action of attention. A regular action of attention involves a sequence of operations (activation of a location, then of a cardinality, and then of a property complex); a WM individual is a working memory device which stores this sequence (see Section 10.6.3). But the representations maintained by an object file are all held simultaneously, and a new WM individual is 'read off' them at single point in time.

Nonetheless, the operation which reads a WM individual off an object file emphasises the idea that the process of monitoring an episode *also functions as an attentional action.* For instance, as discussed in Section **??**, after monitoring a cup-grabbing episode, I end up in a state in which I am (haptically) attending to the target cup. The grab action can also be understood as a deployment of (haptic) attention. Since it involves the movement of a limb, this attentional operation takes much longer than a saccade—but in terms of its end result, it is quite comparable to a saccade. Similarly, monitoring a locomotion action can be thought of as an attentional operation. At the end of the monitoring operation, my attention has been drawn to a new location. Again, the attentional operation is extended in time, because it requires physical movement through space. But its end result is similar to that achieved by making a saccade. In summary, the idea that object files have 'associated WM individuals' can be thought of as encoding the fact that an episode-monitoring process also does service as an attentional operation.

Note that an object file's link to its initial WM individual is a stored association, while its link to its new WM individual is a link to the current perceptual state: the new WM individual is read off perceptual representations. These are quite different operations, which allow the two WM individuals associated with an object file to be clearly distinguished.

11.3.5.2 Object files and the temporal structure of a replayed WM episode

Consider what happens when an observer experiences an event, creates a WM episode representing the event, and then replays this episode in simulation. When the episode is replayed, agent and patient object files will become active at various points. It is interesting to note that both these object files carry information relating to an extended period during action monitoring. During replay, when an object file becomes active, this information becomes available 'out of sequence'. For instance, the first time an object file becomes active, we have access to the WM individual representing the associated object at the end of the action. If the object file represents a tracked group, we might also have information about a change in the numerosity of this group, as just proposed in Section **??**. Similarly, the second time an object file becomes activate, 'late' in the rehearsed sequence, we have access to the WM individual representing the associated object when it was first attended to. These temporal discrepancies are reminiscent of the out-of-sequence signals which form the basis for the account of verb raising given in Section **??** (and just reiterated and extended in Section 11.3.1.1). But note that they have a completely different source. They arise from the storage capacity of object files, not from the storage capacity of WM episodes. Object files and WM episodes are separate, though related, forms of working memory. In Section **??** I will suggest that the storage capacity of object files manifests itself in a different type of long-distance syntactic dependency, connected with the interpretation of quantifiers.

[I also need to say something about **subject-verb agreement** in this chapter. It's not dealt with earlier.]

11.3.5.3 Replay of a WM episode to episodic memory

The other function of replaying a WM episode is to encode the episode in hippocampal long-term memory. (This is probably the primary function of the replay operation—I assume that the mechanism which associates a replayed WM episode with phonological representations evolved later, co-opting the replay operation for a new purpose.) The basic model of replay-to-episodic-memory was given in Section 3.8.1.3. To reiterate: during experience of an episode, the observer evokes a sequence of sensorimotor signals, and stores it in PFC as a WM episode. The sequence is then replayed to the hippocampus, at a speed which allows it to be encoded using LTP. Now that the account of WM episodes has been extended to incorporate WM object representations, this replay process can be described in more detail.

Encoding an episode in LTM must achieve two related things. Firstly, the content of the episode must be stored. As discussed in Section 3.7, the hippocampus stores an episode as a sequence, which echoes the sequence stored in the WM episode. But recall from Section 3.6.4.1 that an episode stored in hippocampal LTM needs to reference LTM individuals, which persist through time, and which can participate in many different episodes. A WM episode represents objects using WM individuals and object files; when it is replayed it must activate the appropriate LTM individuals, so that the hippocampus can build an episode representation which refers to LTM individuals.

Secondly, encoding an episode in LTM must encode the *changes* which the episode brings about. Specifically, any changes to the properties of LTM individuals must be explicitly represented. Recall that the hippocampus encodes the properties of an LTM individual at particular times through the mechanism of context-gated associations. The locations of LTM individuals are stored in context-gated links from LTM individuals to locations (see Section 3.6.4.2). The cardinalities and property complexes associated with LTM individuals, which can also change, are stored in context-gated links from LTM individuals to particular WM individuals (see Section **??**). I suggest that during sensorimotor experience, the hippocampus maintains a representation of 'the current temporal context', and this representation is updated after an event is experienced.

I will now give a slightly more detailed account of how a WM episode is replayed to hippocampal LTM. The main new idea is that the operation of replaying a WM episode to the hippocampus is *synchronised* with the operation of updating the hippocampal representation of the 'current temporal context', in a way which allows changes in the properties of LTM individuals to be encoded at the same time as the content of the event itself is encoded.

I will use our example event *The man grabbed a cup* to illustrate. In working memory, this event is stored as a WM episode which schedules a sequence of three operations: first, activation of an AGENT object file representing the man; next, activation of a PATIENT object file representing the cup; finally, activation of the 'grab' motor programme. As discussed in Section 11.3.5.1, each object file maintains representations of the current properties of its tracked object, which allow a new WM individual (its 'new WM individual') to read off at any time. I will denote these WM individuals WMI_man_{new} and WMI_cup_{new}. Each object file is also associated with an LTM individual, which I will denote *LTMI_man* and *LTMI_cup*. When the WM episode is replayed to the hippocampus, the hippocampus must encode the episode itself, as the sequence *LTMI_man, LTMI_cup, grab*. It must also update the current temporal context, and record any changes to the stative properties of the two LTM individuals.

To begin with, note that the *initial* properties of the man and the cup are encoded in LTM before rehearsal of the WM episode even starts. When the man and the cup are first attended to (in the 'initial phase'), they are each associated with a WM individual. These WM individuals are immediately associated with LTM individuals in the current temporal context; this operation is part of 'recognising' the objects involved (or of creating new LTM individuals if they are not recognised). These LTM individuals are also associated with object files. The rest of the episode is then monitored, and a complete WM episode is created, ready to be replayed.

The first operation in the replayed sequence is defined dually as the operation of activating the AGENT object file, and of activating the WM individual representing the man (see Figure 11.8). Object files are associated with LTM individuals, so this operation also activates *LTMI_man*. The 'man' WM individual is also linked to *LTMI_man* in the intial context, so *LTMI_man* is activated via this route too. The second operation is defined as the activation of the PATIENT object file and the 'cup' WM individual, both of which activate *LTMI_cup* in a similar way. The final operation is the activation of the 'grab' motor programme. I assume that the hippocampus associates the current temporal context with the sequence *LTMI_man, LTMI_man, grab*, so that re-evoking this context will result in the sequence being replayed (see Section **??**).

At the end of a replayed WM episode, the AGENT and PATIENT object files are each reactivated in turn (see Section 11.3.4.1). I now suggest that these reactivations have a special role in encoding the new properties of the agent and patient in LTM. In the scheme I propose, the hippocampal representation of 'the current context' is updated *before* the reactivations take place. When the AGENT object file is reactivated, a special operation links its associated LTM individual (*LTMI_man*) with the WM individual representing the man's current state (WMI_man_{new}), *in the updated context*, and when the PATIENT object file is activated, a similar operation links *LTMI_cup* to WMI_cup_{new} in the updated context. Thus the process of replaying the WM episode drives the LTM encoding of the episode itself, and of the changes which it brings about. Note that the 'special operation' reading out an object file's new WM individual when it is reactivated is an additional stipulation. But the reactivation of the object file itself, and the consequent reactivation of its associated LTM individual, both follow from the existing mechanism.

11.4 The syntax of the DP-clause interface

11.4.1 The role of variables

Basic idea: a verb's arguments are given as variables, which must bind to the referents which are introduced by DPs (see somewhere in Chapter 10). Refer to the 'slot-filling' operation described in Section 6.2.5.3.

11.4.2 Quantifier raising

11.5 A sensorimotor account of the DP-clause interface

11.5.1 A sensorimotor interpretation of variables

11.5.2 A sensorimotor interpretation of quantifier raising: first pass

Also refer forward to Section 12.6.2.

11.6 A processing model of the DP-clause interface

In Chapter 7, I presented an account of how the sensorimotor sequence associated with a reach-to-grasp action is transformed into a sequence of words. The basic idea was that the sequence is stored in a PFC sequence plan (a 'WM episode'), and then replayed, in a mode which allows a sequence of phonological expressions to be read out. 'Read-out' was performed by a special-purpose LF-to-PF mapping network, taking as input a sequence of sensorimotor signals evoked in a range of different sensorimotor and working-memory media, and delivering as output a sequence of words.

The account in Chapter 7 assumed a simple model of objects, in which attending to an object evokes a single sensorimotor representation: say a category like 'man' or 'cup' in IT. In the current chapter I have presented a much more complex model of object representations: the processes involved in attending to objects generate working memory representations of their own, including WM individuals (stored attentional sequences) and object files (encodings of information gathered during sustained tracking operations). The account of how rehearsing a WM episode generates a sequence of words must be considerably extended. The extended account must explain how DPs can be multi-word expressions, with their own internal syntactic structure. And it must provide the basis for a model of the complex syntactic relationship between DPs and clauses.

One thing to note is that working memory now contains several stored sequences, not just one, since attentional operations are associated with stored sequences. While these sequences are all stored simultaneously in working memory, they must be 'read out' as a single sequence of words, because language is a linear medium. My basic proposal is that the operation of 'reading out' a WM episode invokes lower-level operations of 'reading out' individual attentional sequences (i.e. of rehearsing WM individuals) at particular points. I suggest that producing a DP at a certain point in a sentence involves temporarily changing the *control* of the rehearsal operation, to suspend rehearsal of the episode-level sequence while an attentional sequence is rehearsed, and to configure the LF-to-PF mapping network to receive its input from a different source during this process. The LF-to-PF mapping network continues to receive a stream of sensorimotor and working-memory representations, from which a stream of words is generated—however, the *source* of this stream has changed. When rehearsal of the attentional sequence is complete, control is reverted to the episode-level sequence.

What operations trigger the changes in control which allow attentional sequences to be read out? I suggest that object files are the key structures. Recall from Section **??** that the agent and patient object files each become strongly active at two points during the rehearsal of an episode-denoting sequence. I propose that each of these points is an opportunity to enter a special mode of control in which an attentional sequence is rehearsed. A language learner has to learn which of the two opportunities should be taken for each object file, and which should be ignored.

When the activation of an object file does trigger the rehearsal of an attentional sequence, the sequence which is rehearsed must obviously be one which relates to the activated object file. Note that there are two possibilities here. As described in Section 11.3.5.1, an object file is associated with two WM individuals, one depicting the tracked object as it was in the initial state, the other depicting it as it is in the consequent state. Perhaps different types of episode require different ways of depicting participant objects. For instance, maybe the consequent-state WM individual should be used when reporting an episode of object creation (*John made A CAKE*), while the initial-state WM individual should be used when reporting an episode of object destruction (*John smashed A GLASS*). Note that this proposal trades on the fact that object files carry 'out-of-sequence' information, as discussed in Section 11.3.5.2. Thus in *John smashed a glass*, an initial-state WM individual is used to depict the smashed object, even though the referring expression is read out 'late' during rehearsal of the episode. In Section **??** I will argue that quantified noun phrases (e.g. *Most dogs*) result from reading out the information about numerosity change associated with an object file (see Section **??**). This information is also read out out-of-sequence: the two numerosities whose proportion is conveyed are associated with the initial and consequent states.

11.7 Summary

In Section 11.3, I gave an account of how working-memory object representations are involved in monitoring episodes, and subsequently in representing them. Object files are a key construct in this account: my main proposal is that while an episode is being monitored, its participants are represented as object files, which can support changes to the location and/or properties of objects which might take place during monitoring. Object files thus provide the basis for the spatiotemporal continuity of objects during episodes. In addition, once an episode is monitored to completion, object files provide a mechanism for referring to its participants using the properties they have at the start of the episode or those they have at the end. In the next chapter, object files will feature in a sensorimotor account of the 'arguments' of verbs and of their associated thematic roles.

Chapter 12

A sensorimotor interpretation of predication, quantification and relative clauses

12.1 Introduction

In the book so far, I have restricted my attention to a single type of clause: a transitive clause, describing a simple action. In the next two chapters, I will (finally) consider some other clause types. In the current chapter, I will focus on **predicative clauses**: clauses which predicate properties of individuals. Predicative clauses describe states of affairs, rather than events. In English, they are realised using the special verb *be*, often termed the **copula**. Some examples of English predicative clauses are given below:

(12.1) John is happy.

(12.2) Fido is a dog.

The perceptual processes involved in apprehending that an object has a certain property are somewhat distinct from those involved in apprehending an event. And the long-term memory structures used to store properties of objects are also quite distinct from those used to store events. Analogously, the syntax of a predicative clause is somewhat different from that of a transitive clause describing an action. However, in each case there are still some similarities. On the syntactic side, the LF of a predicative clause has a right-branching structure, just like that of a transitive clause. On the sensorimotor side, the processes involved in noticing that an object has a certain property have a strong characteristic sequential structure, just like those involved in experiencing a transitive action. I will argue that the general sensorimotor interpretation of LF proposed in Chapter 5 carries over very neatly to the domain of predication.

In the first part of the chapter, I propose a sensorimotor interpretation of basic predicative clauses. I begin in Section 12.2 by introducing a more detailed model of object categorisation, which includes an account of how individual properties of an object can be apprehended by themselves. In Section 12.3 I introduce a model of **semantic memory**: 'memory for facts', including facts about the properties of individual objects, and facts about the typical properties of objects of particular types. In Section 12.4 I present a model of the syntax of predicative sentences, as usual adopting a Minimalist standpoint. And in Section 12.5 I propose a sensorimotor interpretation of the LF of a basic predicative sentence. This interpretation is consistent with the basic proposal from Chapter 5, that an LF structure describes a replayed sensorimotor sequence.

In the second part of the chapter, I extend the sensorimotor account of predication to cover two other core syntactic topics: quantification and relative clauses. The model of quantification which I develop draws on the account of semantic memory given in Section 12.3. In the basic account, semantic memory is a system of associations between LTM individuals and properties: storing the fact that an object has a particular property consists in strengthening the association between a particular LTM individual and a particular property. I begin in Section 12.6 by giving a preliminary account of quantification in semantic memory, which draws on the idea that a property can function as a cue to retrieve a whole *set* of LTM individuals from semantic memory, rather than just a single individual, and that the operation of property-level IOR can be invoked to inspect the interesting properties of groups of individuals as well as of single individuals. In Section 12.7, I extend the model of semantic memory, in a way which allows episodes to be stored in semantic memory as well as regular perceptual properties. In Section 12.8 I give a more complete model of quantification in semantic memory, which allows for quantified propositions about episodes, and also supports multiply quantified propositions. Finally, in Section 12.9, I make some suggestions about the semantic origins of relative clauses, which extends the new model of semantic memory one step further.

12.2 An extended model of object categorisation: property complexes, categories and competition

In this section I will outline a more detailed model of what is involved in the process of 'object categorisation'. Until now, object categories have been treated as discrete, localised representations; but this is obviously a very big simplification. When an observer visually establishes an object, a complex pattern of activation is evoked in inferotemporal cortex, and throughout visual cortex. What counts as a 'category' representation within these media?

The literature on categorisation is a vast one; some useful entry points are Logothetis and Sheinberg (1996) and more recently Cohen and Lefebvre (2005). I will simply note three of the main findings to emerge in this literature. Firstly, a category is likely to be an assembly of units in different modalities, incorporating representations of shape, colour, visual texture, size and other multimodal properties (see e.g. Pulvermüller, 2001; Bar *et al.*, 2001). These properties become associated together through Hebbian learning, due to their common co-occurrence (see classically Hebb, 1949). Secondly, there is likely to be some internal structure to these category assemblies, with some properties being more closely connected than others (see classically Rosch and Mervis, 1975; Rosch, 1978, and more recently and more recently Huyck, 2007)). Thus there are likely to be some visual stimuli which are especially prototypical of a given category, having only those properties which are most closely associated. Thirdly, it seems likely that when we establish an object, we evoke a set of properties which extends well beyond those which identify it as a certain category, which we might think of as our 'visual concept', or 'imagery', of that specific object (see e.g. Kosslyn *et al.*, 1995; 2006). When we look at two dogs, we can recognise the difference between them, even if we identify them both as dogs. In fact, the set of properties which are evoked when an object is established goes quite some way towards *individuating* the object. Naturally, it does not go all the way. It is still possible that two objects are 'perceptually indistinguishable', and evoke exactly the same complex of properties, and can only be individuated by their location. But it is certainly possible for the perceptual system to deliver distinct representations of two objects even if they have the same dominant type. A model of visual properties must support both a notion of structured categories, and a notion of the rich property complexes that represent individual objects.

The model I propose uses the mechanism of competition between properties, plus a notion of property assemblies, to support these two demands. It is illustrated in Figure 12.1. There are two layers. The **property complex layer** represents all the (visual) properties

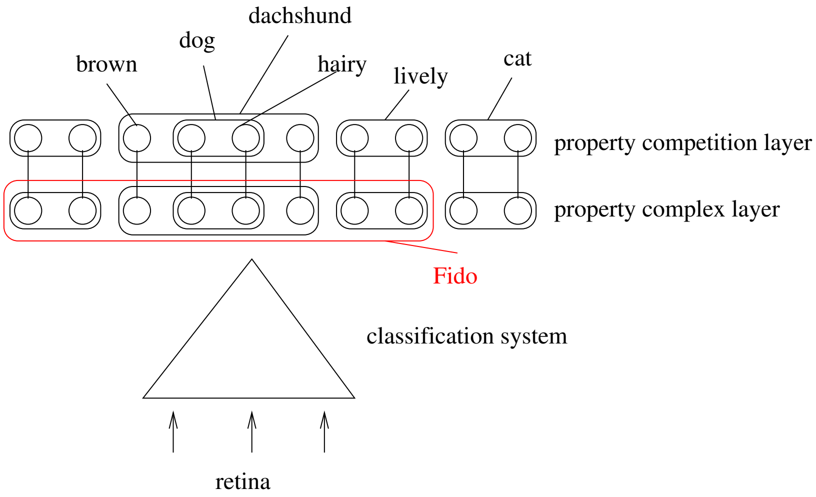


Figure 12.1: A model of object categorisation, including the property complex layer

of an attended object, including the 'shape' properties evoked in IT, but also properties evoked elsewhere in visual cortex. Units in this layer project one-to one to the **property competition layer**. There are inhibitory connections between units in the property competition layer, but not between those in the property complex layer. However, offsetting these inhibitory connections are excitatory connections formed through Hebbian learning. These connections create assemblies of units in both layers, which are depicted by the groupings of units in the figure. Some of the assemblies in the property competition layer represent **object categories**—the kind of things which are denoted by common nouns. There can be hierarchical structure in these categories; for instance, 'dog' is one component of the 'dachshund' category. Other categories are very distinct; for instance 'dog' and 'cat'. There are also assemblies in the property competition layer which do not denote 'complete' categories, which I will term **adjectival properties**, for instance 'hairy', 'brown' and 'lively'. Some of these are components of categories: thus 'brown' is a component of 'dachshund', and 'hairy' is a component of 'dog' (and thus of 'dachshund' also), but they need not be: thus 'lively' is not strongly related to any category. I will refer to the assemblies which can be evoked in the property competition layer collectively as **perceptual properties**. I will discuss categories and adjectival properties in more detail in Sections 12.2.1 and 12.2.2.

Note that the structure of these property assemblies is echoed in the property complex layer. However, the property complex layer also has larger assemblies which identify individual objects, such as Fido, which is a particular combination of properties. I will discuss these assemblies in Section 12.2.3.

12.2.1 Object categories

What constitutes an object category, like 'dog' or 'cat' or 'dachshund'? I propose that an object category is *an assembly which can emerge in the property competition layer when an object is first established.* Assemblies which can achieve this must have two characteristics. Firstly, they must be strongly evoked by visual input. I assume that the assembly 'lively' is not a category because whenever it is evoked, there is always another assembly that is more strongly evoked, and thus dominates in the property competition layer. Secondly, they must be *complete* assemblies. 'Brown' and 'hairy' may be very strongly evoked by visual input, but they form components of larger assemblies, and it is these larger assemblies which become active when objects are first established. The assemblies 'dog' and 'dachshund' are categories, because their component properties have the right combination of being strongly evoked by visual input, and being strongly associated with each other. (There are several reasons why 'dog' might emerge in preference to a more precise category like 'dachshund'. For instance, there might not always be a more precise categorisation: the observer might not recognise certain specific types of dog, or might be classifying a group of dogs of different kinds, in which case the category 'dog' identifies what they all have in common.)

12.2.2 Adjectival properties, and the property-IOR operation

What constitutes an 'adjectival' property, like 'lively' or 'hairy'? These must still be positively defined, as assemblies of co-occurring properties. To define adjectival properties, I will introduce another important operation, which is the *inhibition of the currently dominant property assembly in the property competition layer.* The basic inhibition-of-return circuit has been seen several times: there are inhibitory connections from units in the property competition layer to their corresponding units in the property complex layer. I assume that there is a special operation, which I will call **property-level inhibition-of-return** (**property-level IOR**), which enables these inhibitory connections, and causes the corresponding assembly in the property complex layer to be inhibited. Another assembly will then emerge in the property competition layer.[1] Importantly, the next assembly to emerge can be either another category *or an adjectival property.* Recall that an adjectival property is not able to emerge as a winner in its own right in the property competition layer. However, it can emerge if the dominant assembly (which is always a category) is inhibited. Say we establish an object as a dog, and then inhibit the dominant 'dog' category in the property complex layer. Let us say that the established dog was unusually hairy (for a dog). I assume that the property assembly 'hairy' could then be the new winning assembly in the property competition layer.

Note that the assembly which emerges in the property competition layer after property-level IOR could also be another category. For instance, after inhibiting 'dog' in the above case, it may be that the category 'dachshund' emerges as the new winning assembly. This category might still include properties from the inhibited 'dog' category, because the 'dog' category is part of the 'dachshund' category, and the dachshund-specific properties will try to activate them on this basis. But these properties will be de-emphasised in relation to the dachshund-specific properties, because of the inhibition they are receiving.

What is the purpose of property-level IOR? Essentially, it is an operation which allows an observer to 'attend' to specific properties of an established object. Because of the inhibitory nature of the operation, the properties which are attended to will be those which maximally distinguish the object from the prototypical instance of its category—i.e. which distinguish it most strongly from the objects which it 'most resembles'. Note that the operation is well-suited to support the refinement or elaboration of the observer's system of categories and properties, and can underlie the creation of new subtypes of a given type, or new properties which can be possessed by objects of different types.[2]

To anticipate Section 12.5: I will argue that the property-level IOR operation is the denotation (or rather, one of the denotations) of the verb *be*. Stative sentences which describe the properties of an object (e.g. *X is hairy*), or which classify an object (e.g. *X is a dachshund*), will be understood as descriptions of property-level IOR operations. The noun phrase which is introduced by *be* is typically analysed as denoting a property, like an adjective, rather than as fulfilling a 'referential' function; the notion of property-level IOR will play a useful role in accounting for this property-denoting use of noun phrases.

It is useful to consider property-level IOR in relation to the other forms of self-inhibition

---

[1] Another way of modelling this self-inhibition operation would be to use a single layer of units, and assume some form of habituation or fatigue which inhibits units. This type of inhibition can be implemented in an extension to a regular Hopfield network, allowing it to cycle through a sequence of attractor states (see e.g. Treves, 2005). However, the inhibition operation in such cases is automatic—habituation cannot be 'disabled'. I am using two layers to allow a similar effect to be achieved via an explicit cognitive operation, which can be 'selected' among a set of other possible operations.

[2] Is this idea of property-level IOR already out there? If so, I need to cite the original idea.

which we have introduced so far. Regular 'spatial' IOR consists in shifting attention to a different spatial location (see Section 2.4.5). Another form of IOR is to inhibit the spatial frequency currently being used to provide input to the object categorisation system (see Section 10.5.4). Note that the attended location is not changed here; the current location is simply reparsed at a different spatial frequency by the categorisation system, allowing for the posssibility of 'group classification'. Another form of IOR is to change the saliency map which is currently being used to control shifts in visual attention (see Section 10.8.3.1). This allows us to 'zoom in' to the currently attended object, by adopting a saliency map of locations within this object, or to 'pull back', and treat the whole of the currently active saliency map as a single object in a larger saliency map. By comparison, with these operations, property-level IOR is just another self-inhibition circuit in the visual system. The fact that the same basic circuit has multiple different uses within the system is an attractively parsimonious feature, especially if each circuit can be associated with a particular linguistic phenomenon.

12.2.3 Individual objects and assemblies in the property complex layer

Due to its competitive nature, the property competition layer only ever holds a partial representation of an individual object's properties. However, a much richer representation is held in the property complex layer. As already mentioned, I assume that assemblies in this layer can go some way towards individuating individual objects. In this section, I will discuss these individual-denoting assemblies in some more detail.

For one thing, assemblies of properties in the property complex layer might provide a way for representing the ensemble of an object's current properties, even if these are not all perceivable simultaneously. For instance, such an assembly may be able to represent perceptual properties of the front and the back of an object simultaneously, even if these cannot be perceived simultaneously. Recall from Section 11.2 that Kahneman and Treisman's model of object files required a mechanism for representing the nonvisible properties of an object currently in view. The property complex layer can be thought of as providing this mechanism. Of course, it is very important to distinguish between changes in an object's manifest properties which are just due to perceptual actions, and changes which result from actual changes in the object. I will assume that an assembly in the property complex layer can group properties of an object which are *perceived* at successive moments in time, provided its properties do not *actually* change during the interval between these moments. Distinguishing between 'actual change' and purely 'epistemic change' presumably requires reference to the observer's perceptual operations (understood in a wide sense, encompassing movements and directions of attention); I will not attempt to go into details. But note that any 'actual change' in the object constitutes an *event*, and should be registered using the kind of change-detection mechanisms discussed in Section 11.3.

If a property complex can represent non-visible properties of an attended object, then we can see how it is that a rich representation of the properties of the object could be evoked by completion, if a sufficiently distinguishing set of properties is activated bottom-up. The completion operation would effectively implement a form of 'visual inference' about an attended individual object's hidden properties.

12.3 Semantic memory: memory for the properties of objects

In Section 11.3 I outlined an account of the relationship between object and episode representations. This account focussed on working memory, describing how working memory object representations (object files and WM individuals) participate in WM episode representations. In this section I will argue that objects and episodes are also related together in long-term memory, as part of the so-called 'semantic memory' system. Semantic memory is typically very broadly defined, encompassing several different types of declarative knowledge. I will argue that one variety of semantic memory is implemented through direct long-term associations between LTM objects and property complexes.

12.3.1 Episodic and semantic memory

As briefly mentioned in Section 3.6, there are traditionally thought to be two forms of declarative long-term memory: 'episodic memory', which holds sequences of episodes witnessed by the agent, and 'semantic memory', which holds 'facts about objects' (Tulving, 1972).

Episodic memories are tied to specific spatial and temporal contexts, and thus encode token events. Evoking a context enables recall of the event which happened in that context, which in turn allows an updated context to be activated and the next event to be recalled—thus episodic memories have a temporal sequential structure.

Semantic memory is memory for 'facts about objects'. It is normally understood quite broadly; for instance, it encompasses facts about the perceptual properties of objects (such as 'X is a dog', or 'X is brown'), as well as more derived 'encyclopaedic' properties (e.g. 'X is the capital of Sweden'). Items stored in semantic memory often generalise over spatiotemporal contexts, rather than being tied to specific contexts. Rather than being indexed to contexts, they are often understood as being indexed to objects (citations needed here). Thus semantic memory allows us to activate an object representation and retrieve 'facts about' this object.

Importantly, semantic memory also permits generalisations *over objects* to be expressed. There are two interesting kinds of fact which generalise over objects. **Generic** facts identify the typical properties of individuals of particular types: for instance 'birds can (typically) fly'. Another group of facts, which we might call **finitely quantified** facts, identify properties which are shared by particular groups of individuals: for instance 'both of my children like cheese'. Both types of generalisation can be restricted to given spatial or temporal contexts, provided these are not specific enough to individuate a single fact. Thus it is fine to say 'last Tuesday people typically woke up early' (provided the context identifies many people), or ‘last Tuesday John was normally grumpy with his clients’ (provided the context allows there to be many clients, and thus many opportunities for him to be grumpy), but it is odd to say ‘last Tuesday John typically woke up early’ (because John is unlikely to wake up sufficiently many times in the selected context to warrant a generalisation about the event).

12.3.2 A simple model of semantic memory

In this section I will outline a simple account of semantic memory. The account has the same organisation as the account of episodic memory given in Chapter 3. I will begin in Section 12.3.2.1 by discussing the structure of semantic memory; in Sections 12.3.2.2 and 12.3.2.3 I will discuss the processes whereby items are stored in and retrieved from semantic memory.

12.3.2.1 The structure of semantic memory

Recall that semantic memory is a collection of facts about objects. I propose thinking about semantic memory in general as a set of direct, long-term associations between LTM individuals and a variety of other representations which encode ‘properties of objects’.

To begin with, I will focus on simple perceptual properties, of the kind which are directly evoked during perceptual categorisation. The relevant notion of property was introduced (for the visual modality) in Section 12.2, in the extended model of visual categorisation given there. In that section I introduced two layers of visual properties: a **property complex layer**, representing a rich ensemble of an established object’s properties, and a **property competition layer**, representing the same properties, but in a mode enforcing the competitive selection of a single assembly of properties. My basic proposal is that semantic memory consists of direct associations between LTM individuals and units in the property complex layer. In fact, we have already posited the existence of such associations: in the revised model of LTM individuals given in Section 10.8.3.4, each LTM individual is assumed to be linked to a particular set of units in the property complex layer. The idea is illustrated in more detail in Figure 12.2. The main new element in the model of semantic memory which I will introduce here concerns how these associations are created, or strengthened.

12.3.2.2 Predicative propositions and the formation of semantic memories

The associations between LTM individuals and property complexes which were discussed in Section 10.8.3.4 are all created simultaneously. When an object is established, a property complex is activated, and the whole complex is then associated with a given LTM individual. However, there are also situations where *particular* properties or categories are registered: for instance, when an observer registers that ‘X is a dachshund’, or ‘X is brown’: these are the cases I will focus on here. There is evidence that ‘noting’ a property helps encode the property in memory: for instance, if a subject’s attention is drawn to a

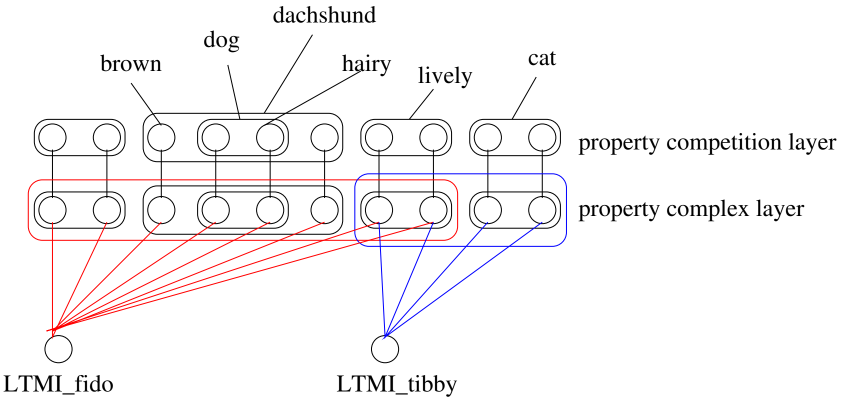


Figure 12.2: The structure of semantic memory: associations between LTM individuals and property complexes

property, this property is more readily recalled (citation needed). We need a model of how 'specific properties' are encoded in semantic memory.

Recall from Section 12.2 that the function of the property competition layer is to isolate specific properties of an individual from a complete property complex. The kinds of assembly which can appear in the property competition layer when an object is first established were termed 'categories'; the others, which can only be activated in the property competition layer after the currently dominant property assembly is inhibited, were termed 'adjectival properties'. The operation of inhibiting the currently dominant property assembly, or 'property-level IOR', was invoked to allow a method of attending to particular properties of an object—specifically, those properties which most strongly distinguish it from other objects of its type. I suggest that the property-level IOR operation has side-effects in the semantic memory system, strengthening the association between an LTM individual and the newly-evoked properties, in a way which emphasises these properties in the individual's property complex.

The observation that a particular individual has a certain property (e.g. 'X is a dachshund', or 'X is brown') is a proposition—something which requires a whole clause to be expressed linguistically, just like the cup-grabbing episode. What are the cognitive processes which underlie these propositions? I want to retain the idea from Chapter 2, that 'propositions' are sequences of cognitive operations. Consider a very 'concrete' item in semantic memory: for instance, an association between an LTM individual (say a dog) and a colour property (say 'brown'). I assume that the process involved in registering that 'the dog is brown' has internal sequential structure. Specifically, I assume that it comprises three operations: firstly an action of attention to the dog; secondly the execution of the property-level IOR operation, which evokes the most distinctive perceptual properties of the dog, and finally, the evocation of one particular property, 'brown'.

This sequence of representations must somehow have the effect of creating or strengthening associations in semantic memory. If we were encoding an action or event in episodic memory, we would store the sequence in working memory, and then replay it for storage in the hippocampus. In the case of a sequence involving property-level IOR, I suggest that we create a working memory sequence, just as for an episode, but that when this sequence is replayed, the IOR operation functions to initiate a different form of storage, which involves strengthening associations in semantic memory, rather than adding an item to episodic memory.[3]

The first element in the replayed sequence is the activation of an LTM individual. When the second element, the IOR operation, is detected, I suggest this functions to *retain* the active LTM individual, and to move into a mode where any subsequently evoked property will be associated with this individual. The final element in the replayed sequence is a property, which will thus be associated with the retained individual.

Note that semantic memory consists of associations between LTM individuals and units in the property complex layer. 'Specific properties' are assemblies in the property competition layer: we do not want these assemblies to be directly linked to LTM individuals. Rather, we want activation in these units to enable associations to be made between LTM individuals and the corresponding units in the property complex layer. A network which allows this type of strengthening is shown in Figure 12.3. (The red 'gate' connections should be thought of as enabling *plasticity* in specific synapses, as well as just gating information flow.) When a unit is active in the property competition layer, it enables plasticity in the synapses of its associated unit in the property complex layer. Thus when a given LTM individual is active, and a particular pattern is active in the property competition layer, associations are strengthened between the LTM individual and the associated pattern in the property complex layer. The effect of this is to emphasise a particular property in the property complex related to this individual.

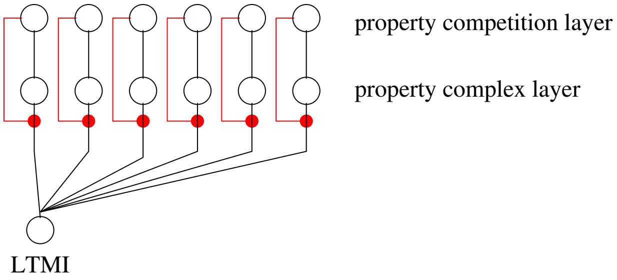


Figure 12.3:

12.3.2.3 Retrieval of semantic memories

If the property-level IOR operation is what causes semantic memories to be stored, a question then arises as to how they are retrieved. As in Section 3.8.2, it is useful to distinguish

---

[3] More generally, I would like to suggest that the distinction between episodic and semantic memory is at the root of the distinctions between 'states' and 'events', which semanticists typically distinguish by means of ontological primitives.

between 'recognition' and 'recall' memory retrieval tasks. Recognition involves presenting a complete proposition as a memory cue, and determining whether this proposition was stored in memory. Recall consists in presenting a partial proposition as a cue, and using memory to complete the proposition. I will focus on recall in this section.

There are two possible kinds of recall for semantic memory. One consists in presenting a property as a memory cue, and activating LTM individuals which have this property. The other consists in presenting an LTM individual as a cue, and retrieving the properties associated with this individual. I will consider these cases in the remainder of this section.

Properties as memory cues: activating groups of LTM individuals Semantic memory stores many-to-many associations between properties and LTM individuals. This means that if we activate a particular property as a memory cue, we will activate *all* the LTM individuals which have that property. 'Activating a property' consists in activating an assembly in the property competition layer. This also gates open the associations between the corresponding assembly in the property complex layer and the set of LTM individuals. If we assume that activation flows from property competition units to property complex units when a property is presented as a memory cue, then activating a property will activate all the LTM individuals which have this property as part of their property complex.

The retrieval of a group of individuals which all share some property is another form of 'attention to a group of individuals', which in some sense is a memory analogue of perceptual 'group classification'. In each case, the observer establishes a group of objects, which are all of the same type. In each case, therefore, we can represent the established group with a plural noun (e.g. *dogs*). There are some apparent disanalogies, though. In group classification, the classified group is associated with a region in a saliency map, and a sub-saliency map is created to represent this region, whose locations represent individuals in the group. This saliency map allows individual members of the group to be attended to in turn, and can be used to obtain a measure of the cardinality or numerosity of the group.

Clearly in a situation where a group of objects is retrieved from semantic memory, there is no concrete spatial component to the state of attending to a group. However, it is often suggested that something quite similar to a spatial representation is used to represent groups of individuals recalled from semantic memory, supporting operations such as the partitioning of individuals into different groups according to their properties, or iterated attention to each individual in turn. (The idea of placing individuals recalled from memory into 'spaces' is particularly explicit in Johnson-Laird's (1983) theory of 'mental models'.) I will assume that when a property is used to recall a group of LTM individuals from semantic memory, the activated individuals are associated with points in a saliency map, which can be explored in the usual way by inhibition of return, and whose numerosity can be directly assessed.

[I will also assume that the group of retrieved individuals is associated with a WM individual, and perhaps with an object file, but I'm not sure how that story goes yet.]

LTM individuals as memory cues An LTM individual must also be able to function as a memory cue. In this scenario, the observer activates an LTM individual, and thereby retrieves an associated property complex.

To begin with: how can an LTM individual become active? One method is through perception: if an object is established, the perceptually evoked property complex and location can be sufficient to activate an LTM individual, as described in Section **??**. In this case, 'retrieving' the property complex associated with the object is a matter of completing the cue pattern. However, LTM individuals can also become active in memory mode. For instance, an event in episodic memory is stored as a sequence of representations, some of which are activations of LTM individuals (see Section **??**). Similarly, we can use properties as cues to activate a group of LTM individuals, as just described. Assuming we create an abstract 'space' to support selective attention to members of this group, as proposed above, we can also activate LTM individuals by cycling through a set of individuals activated by querying semantic memory with a property cue. In both cases, when an LTM individual is activated, its associated property complex will be activated.

Note again that there are two ways of activating an LTM individual: one is perceptually, and one is via a memory operation. In each case, the observer ends up with an active LTM individual, and an active property complex. And in each case, an activated LTM individual is associated with a location in a saliency map, whether this represents a point in physical space, or a point in an abstract space of items retrieved from memory.

Having established an LTM individual, I suggest that its properties can be attended to selectively, just as they can in perceptual mode, by executing the property-level IOR operation. The distinguishing properties which this operation reveals are most likely to be properties which were noted at the time it was encountered perceptually. But it is also possible to recall distinguishing properties which were not noted perceptually, even though these will probably not be so salient in the property complex.

12.4 The syntax of predication

This needs to draw on Moro and Heycock.

12.5 A sensorimotor interpretation of predication and properties

Say something in here about the verb *be* (the copula) denoting property-level IOR.

12.6 Quantification and the semantic memory system: a preliminary account

As mentioned in Section 12.3.1, one of the key characteristics of semantic memory is that it supports generalisations over individuals. In this section, I will give a preliminary account of how quantified propositions are represented in the semantic memory system. This account will deal with quantified sentences expressing that a group of individuals have a particular property: for instance, *Most dogs are brown* or *Many dogs are black*. A more detailed account of quantification is given in Section 12.8.

12.6.1 The basic idea

[This section needs to be fleshed out.]

I suggest that underlying the quantified proposition 'Many dogs are black' is a sequence of two operations accessing semantic memory. The first operation is the presentation of a property ('dog') as a cue to semantic memory, which results in the establishment of a group of LTM individuals. As a result of this operation, each of the active LTM individuals activates its associated property complex in the property complex layer. The resulting pattern of activation is a reflection of the *combined properties* of the activated set of dogs. From these combined properties, a single property is selected in the property competition layer. It is to be expected that the first property to be selected will be the category 'dog', because it was on the basis of this property that the individuals were selected.

The second operation is the inhibition of this dominant property, i.e. property-level IOR. This is the crucial operation for quantification. At this point, the new property which emerges in the property competition layer will be a reflection of the idiosyncratic properties of the activated individuals *seen as a group*. There are two separate reasons why a property may compete strongly after inhibition of the dominant 'dog' category. One is that there are several LTM individuals which have a particular idiosyncratic property. (The more LTM individuals which have it, the more strongly it will be activated.) Another is that the property may be highly idiosyncratic. (All other things being equal, properties which are very unusual for dogs will be more more strongly activated, or at least less strongly inhibited, than those which are less unusual.) Thus the winning property may be possessed by a large number—possibly all—of the activated LTM individuals, or it may be possessed by a small number of them, if it happens to be very unusual.

Note that quantified propositions provide some information about the number of LTM individuals which possess the winning property. ('All the Xs are Y' indicates that they all do; 'Some of the Xs are Y' indicates (roughly speaking) that a small number do, and so on. I will not attempt to give definitions of the cardinalities or proportions which are picked out by the different quantifiers; however, it is important to consider the mechanism which allows the cardinalities of the initially established set of LTM individuals, and of the subset of these individuals which possess the winning property, to be registered. The mechanism I propose draws on the idea that establishing a set of LTM individuals by querying semantic memory with a proposition has many similarities with perceptual group classification: in each case, a saliency map is created to represent the items in the established set. As suggested in Section 12.3.2.3, a saliency map can be used to generate a measure of cardinality or numerosity. I further suggest that after property-level IOR occurs, the number of active LTM individuals reduces to those which have the new winning property, and that a new sub-saliency map is created to represent the remaining active individuals. Different quantifiers can then be read off the ratio of the cardinalities/numerosities of the two saliency maps, or off the absolute cardinality/numerosity of the second map, or perhaps off some combination of these two measures.

12.6.2 Object files and the non-locality of quantified DPs

Recall from Section 11.4.2 that a quantified DP contains information which scopes over the whole clause in which it appears. For instance, in the clause *Many dogs are black*, the quantifying determiner *many* tells us not just about the set of dogs, but about the set of dogs which are black. In Section!11.5.2 I proposed that the nonlocality of quantified DPs relates to the fact that each DP is associated with a unique *object file*, which tracks a particular individual throughout an episode. In Chapter 11, the main purpose of object files was to maintain the identity of single objects as they participate in episodes, potentially changing their location or intrinsic properties. I now suggest that object files also have a role in quantified sentences, where DPs represent groups of individuals activated in semantic memory, rather than objects tracked in temporally evolving real-world contexts.

My key proposal is that the reduction in the size of the attended group *is registered in the object file representing the group of dogs.* An object file tracks an object during the perception of an episode, and is invariant to changes in its location or properties. An object file can also track an attended *group*. In this case, I suggest it is also invariant to changes in the size of the group—i.e. that it remains associated with the group if its size is reduced. Recall from Chapter 11 that an object file is associated with two WM individuals, one at the start of the action and one at the end. I suggest that a WM individual representing a group encodes the size of the group, using the pre-numerical **numerosity** representation discussed in Section 10.5.4. (Recall that a WM individual is a planned attentional sequence: to establish a location, then a cardinality and then a template. I suggest that the first operation in this sequence, reference to a location, can include a measure of the numerosity of the group established at that location.) As discussed in Chapter 11, the mechanism for 'reading out' an object file during sentence generation can draw on both of its associated WM individuals. For a 'quantified' object file, whose two WM individuals represent different numerosities, I suggest that there are two ways of 'reading out' the numerosity of a DP. One way is to read out the *proportion* of the numerosities of the initial and final WM individuals. The other is to read out the absolute value of the final WM individual. The former method results in 'proportional' quantifying determiners (e.g. *most*, *all*, *many*) while the latter results in 'cardinal' determiners (e.g.

some, *one*, *no*).[4] Thus in our example *Most dogs are brown*, the object file associated with the set of dogs holds a record of how establishing the predicate 'brown' diminished the numerosity of the set of dogs. When the sentence is rehearsed, the object file's information about the relative numerosity of the two sets is 'read out' at a single point, even though it was gathered over an extended period of time which is basically identical to the time taken to monitor the episode reported by the sentence.[5]

The treatment of quantifiers in the above account hinges on the relationship between object files and the WM episodes in which they feature. An object file is linked to two WM individuals—one recording the initial establishment of an object or group, and one describing how to re-establish this object/group when the associated episode is complete. In other words, object files and WM episodes are updated in parallel as the episode is monitored, and *collectively* provide information about the processes which occur while it is monitored. However, when a WM episode is replayed and transformed into a sequence of words, there are only certain specific points at which the object file is 'triggered', allowing its associated WM individuals to be rehearsed. While these points correspond to particular points in the monitoring of the episode, the information *read out* from the object file spans a wider time period, which might begin well before it and end well after it. In syntactic terms, we can say that a quantified DP holds 'non-local' information, which relates to points elsewhere in the structure of the clause. In a standard syntactic model, this non-locality is modelled by the mechanism of quantifier raising, a type of movement which is distinct both from the raising of DPs to Case-assigning positions and from head-to-head movement. In the sensorimotor account, the non-locality associated with quantifiers also has a distinct origin. The non-locality represented by DP raising reflects processes of reattention to agent and patient. The non-locality represented by head-to-head movement reflects tonically active representations in sequence plans. The non-locality in a quantified DP reflects the fact that an object file represents processing over an extended period of time during the monitoring of an episode, but its associated WM individuals must be *rehearsed* at a single point.

It is interesting to compare this 'sensorimotor' account of non-locality in quantified DPs to the standard account of quantifier raising. Figure 12.4 shows the QR analysis of our example, *Most dogs are brown.* The top IP projection in this structure introduces the raised DP, in which the noun *dogs* denotes the initially-established set of dogs. The lower IP denotes the set of brown things. The movement allows *most* to denote a function over these two sets. The two stages of the perceptual process underlying the sentence are roughly echoed in the LF structure of the sentence after quantifier raising. We could

---

[4] Of course, the exact proportions for the 'vague' quantifiers are hard to determine. One useful suggestion is that quantifiers can denote *deviations from expected proportions* (Ferndando and Kamp, 1996). Since the account I just proposed makes reference to typicality relations between predicates, it is potentially consistent with this idea.

[5] One thing missing from this account: syntacticians often assume that cardinal determiners (e.g. *a*, *two*, *no*) originate at Num/PD and raise to SD when they need to refer (e.g. outside existential contexts). I have a story about the distinction between singular and plural at NumP, but I don't have a story about how 'counting' can happen at NumP.

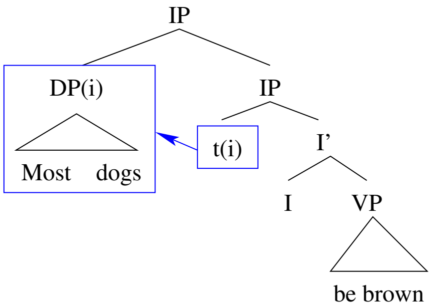


Figure 12.4: The LF of *Most dogs are brown* after quantifier raising

interpret the constituent holding the raised DP as denoting the process of establishing the set of dogs, and the nested IP as denoting the subsequent process of establishing the set of brown things. However, the position of the quantifier *most* is at odds with our standard temporal interpretation of right-branching LF structure. It denotes the reduction of the cardinality of the set of dogs after establishment of the 'brown' predicate, which is the very last perceptual operation. If LF denotes a sensorimotor sequence, *most* should appear at the bottom of the tree, not at the top. Yet if it were in this lower position, its association with the set of dogs would not be explicit.

In other words, my account of quantifiers cannot be thought of as a 'sensorimotor reinterpretation of' quantifier raising. It has to be thought of as an *alternative* account of the syntax of quantifiers. (There are many existing accounts of how quantified DPs can be interpreted without movement, which typically assume some form of memory associated with DPs—see e.g. Cooper's (1983) model of 'quantifier storage'. My account could perhaps be compared to these.) Note that quantifier raising is an unusual form of movement in any case. For one thing, the IP projection created when the DP is raised is not a normal X-bar structure: it does not have a head. Moreover, quantifier raising is never explicit at PF: it is always assumed to occur after spellout. So there are some advantages to a non-QR account of quantifiers.

12.7 Semantic memory for episodes

Events can also feature in semantic memory. For instance, we can express generalisations about the participation of objects in events: these can be generic (e.g. 'Italians typically put sugar in their coffee') or finitely quantified (e.g. 'Everyone in this office graduated from Harvard'). These statements each include the representation of an event—'putting' or 'graduating'—but they are hard to understand as part of episodic memory. For instance, 'Everyone in this office graduated from Harvard' probably summarises a set of events that happened at widely different times. Rather they are facts about groups of people, or types of people.

It is interesting to consider generalisations of our cup-grabbing event: for instance, ‘every student grabbed a cup’, or ‘John grabbed most cups on the table’. Clearly, these generalisations make reference to a grabbing episode. So far, we have considered how a single token cup-grabbing episode is stored in working memory and in episodic long-term memory. Presumably coming to *know* these generalisations requires experiencing actual episodes. How does the representation of a generalisation about the episode relate to the representation of an individual episode? And how is the generalisation derived from the individual episodes?

I propose that WM episodes play a role in representing generalisations over episodes, as well as individual episodes, and that WM episodes play a key role in explaining how generalisations over episodes can be formed from experience of individual episodes. My proposal will draw on the idea that object files can function as *variables* in WM episode representations, as discussed in Section 11.3.4. That idea can be summarised as follows. Object files carry information about the role a tracked object plays in an episode being experienced, while abstracting away from all the other properties of the tracked object. This abstraction allows it to represent the continuity of objects as they participate in events which *change* their properties. But it could also provide a means for WM episodes to participate in representations of generalisations over groups of objects.

12.7.1 Abstracted WM episodes and semantic memory

So far I have focussed on simple perceptual properties. But as we have already seen, properties can also make reference to episodes. In this section I will outline an idea about how arbitrary WM episodes can feature in semantic memory.

The basic proposal is that LTM individuals can be associated directly with *WM episodes* in semantic memory as well as with perceptual properties. For instance, after a cup-grabbing episode is experienced, the LTM individual denoting the agent could be linked directly to the WM episode, so that activating the WM episode can evoke the LTM individual. This would encode the agent’s participation in the episode as a ‘fact’ about the agent. The way the cup-grabbing episode is stored in episodic memory is quite different: it is associated with a temporal context, and when it is rehearsed, it activates the subsequent temporal context.

Linking a LTM individual to a WM episode affords various opportunities for abstracting away from details of the episode, to support the expression of generalisations over individuals. Most obviously, the episode does not need to include a representation of the individual to which it is linked. Say we are linking the WM episode ‘John grabbed a cup’ to the LTM individual *John*. Rather than linking the LTM individual to a fully specified representation of the episode, we can link it to a representation which leaves the agent unspecified. Recall from Section 11.3.4.2 that WM episodes represent the participants in an event by object files, as well as by WM individuals. As shown in Figure 11.9(a), repeated below as Figure 12.5, referring to the agent participant *just as the agent* is akin to referring to it as an unbound variable. The operation of linking this WM episode to the LTM individual *John* is akin to the operation of *binding* the variable to a particular individual.

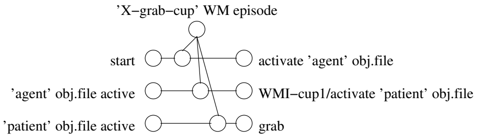


Figure 12.5: WM episode representing 'X grabbed a cup'

Note that the WM episode 'X grabbed a cup' can be linked independently to many LTM individuals, and thus supports generalisations over LTM individuals. In this sense it is similar to the perceptual properties which were discussed in Section 12.3.2. In fact, the operation of turning an episode into a property by referring to one of its participants as an unbound variable will be familiar to natural language semanticists, who term the operation 'abstraction'. To echo this terminology, I will refer to a WM episode containing exactly one 'bare' object file reference (i.e. one free variable) as an **abstracted WM episode**. (To refer to perceptual properties associated with object categories I will keep using the terms 'categories' or 'types'.)

12.7.2 Reference to individuals in abstracted WM episodes

One issue left unresolved in the previous section is how the *remaining* participants in an abstracted WM episode are represented. These participants are referred to by WM individuals as well as by object files. But note that object files are only transiently associated with specific LTM individuals, and WM individuals can be associated with many LTM individuals. If a WM episode is retrieved during semantic memory access, long after the episode was experienced, it may well be impossible to recover the LTM individuals which participated in it. I suggest that there are several methods by which individuals can be referred to in an abstracted WM episode. These will be discussed in turn below.

12.7.2.1 'Existential quantification' in abstracted WM episodes

One option is to continue to refer to a participant in an abstracted WM episode using an object file and a WM individual, even though the working-memory associations between these constructs and a particular world object will be lost. What form of reference is achieved by an 'anonymous' object file and an associated WM individual in an abstracted WM episode? I suggest that this form of reference amounts to what logicians call 'existential quantification'. For instance, the abstracted WM episode shown in Figure 12.5 represents 'X grabbed *some* cup', rather than 'X grabbed *a particular* cup'. The object file functions somewhat like an existential quantifier, introducing a variable referent. It testifies that some object was established, but says nothing more about it. The WM individual predicates some properties of this referent. These properties can be more or less specific, but they may well fall short of identifying an individual.

The fact that abstracted WM episodes can be associated with many LTM individuals, together with with their ability to represent participants using this 'nonspecific' or 'quantifying' form of reference, allow some interesting generalisations to be expressed. Say an observer watches John grab one cup, and then Bill grab another. Each event can be represented by the same abstracted WM episode, whose agent is a bare object file, and whose patient is a nonspecific WM individual representing a cup—just as shown in Figure 12.5. This episode can then be linked in semantic memory to the LTM individual representing John, and also to the LTM individual representing Bill, even though they grabbed different cups. Linguistically, this generalisation might be expressed as *John and Bill grabbed a cup*, with *a cup* understood in its 'nonspecific' or 'quantified' sense. In first-order logic, this nonspecific sense is achieved by introducing an existential quantifier within the scope of another quantifier (for instance 'both', or 'all'), to allow different bindings for the cup for the two agents. The fact that WM individuals in abstracted WM episodes have this nonspecific sense further emphasises that they play a role akin to existential quantifiers.

Of course, generalisations can also refer to individuals. If both John and Bill chased Sally, I can frame the generalisation 'John and Bill chased Sally', as well as 'John and Bill chased a girl'. Since WM episodes do not make direct reference to LTM individuals, the former generalisation cannot yet be expressed within the scheme I have introduced. To include it, I must allow the abstract WM episodes which make direct reference to LTM individuals, as shown in Figure 12.6. In order to allow this new form of reference to individuals in WM episodes, we have to allow that the observer can *prepare* the action of activating an arbitrary LTM individual, as part of a larger prepared sequence. This will clearly involve some additional machinery—minimally, a small number of 'pathway units' for each LTM individual. But given that pathways are represented in terms of object files (see Section 11.3.4.1), the amount of additional machinery required is not excessive.

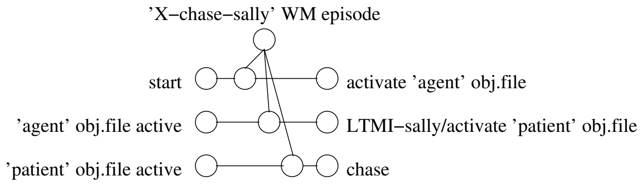


Figure 12.6: An abstracted WM episode making direct reference to a LTM individual

12.7.2.2 Consolidation and the formation of abstracted WM episodes

When are LTM individuals linked to abstracted WM episodes? The most obvious time is immediately after an episode is experienced. Presumably it can sometimes happen at this point, if the observer happens to reflect in a certain way on the episode. However, this is not the only time it could occur. It may be that the creation of abstracted WM episodes of appropriate generality happens more systematically, during operations which systematise an observer's world knowledge. The creation of generalisations is often hypothesised to be one of the functions of the 'consolidation' operation, which transfers episodic memories stored in the hippocampus into more permanent cortical storage. Many theorists have suggested that the cortex stores episodes in a format which supports the extraction of generalisations (see e.g. McClelland *et al.*, 1995). Perhaps the process of creating cortical memories from hippocampal ones provides an opportunity to represent episodes as items in semantic memory.

12.7.2.3 Competition between episode-level properties

The 'property complex' associated with an LTM individual now contains a mixture of perceptual properties and WM episodes. We can therefore envisage the creation of assemblies representing types of object which include a mixture of both kinds of property. For instance, the assembly which represents the category 'dog' might include various perceptual properties, but also the abstracted WM episode 'X chases a stick': an episode in which dogs frequently participate. In this context, it should be possible for the property evoked by property-level IOR to be an episode-level property as well as a perceptual one. For instance, the most unusual fact about a given dog might be that it frequently drinks tea: in this case, we should expect the abstracted episode X drinks tea to be selected by property-level IOR. In order to support this, we must envisage that each abstracted WM episode is represented twice, once in the property complex layer and once in the property competition layer: abstracted WM episodes can then compete against perceptual properties to be evoked by property-level IOR.

12.8 An extended model of quantification in the semantic memory system

In this section, I give an extended account of quantification, drawing on the idea of episode-level properties just outlined in Section 12.7. We saw in Section 12.7.2.1 how an abstracted WM episode can be linked to several different LTM individuals, to support statements about groups of individuals (e.g. 'John and Bill grabbed a cup'). However, to give an account of 'quantified' statements, such as 'Many men grabbed a cup', or 'John grabbed many cups', some additional machinery needs to be introduced. I will give an account of simple quantification in Section **??**, and extend it to multiple quantification in Section 12.8.0.5. [6]

12.8.0.4 Simple quantified propositions containing episodes

[This needs rewritten]

---

[6] Somewhere in here I should refer to Rips (1975) as one of the original proposals that quantification relates to semantic memory.

12.8.0.5 Multiply quantified propositions

It appears possible to express propositions in semantic memory which record multiple generalisation operations. For instance, I can report that *Many boys chased every girl.* As is well known, sentences which report multiple generalisations display **quantifier scope ambiguity**: for instance, the above sentence could mean that there is a particular group of boys such that each boy in that group chased every girl; it could also mean that for every girl it is true that many boys chased that girl (but it could be a different set of boys for each girl).

How can we account for multiply quantified propositions in the semantic memory scheme we are developing? My basic proposal is that quantified propositions can *themselves* be expressed as WM episodes, which can themselves be treated as properties, and indexed to LTM individuals in their own right. For instance, say an observer watches a series of events in which a boy chases a girl. He first watches Jim chase a number of girls: Sue, Mary and so on. In each case, he chooses to record the chasing episode as a property of the chased girl, by linking LTM$_{sue}$ to *Jim chased X*, LTM$_{mary}$ to *Jim chased X* and so on (see Figure 12.7(a)). Next, the observer happens to observe a collective property of the

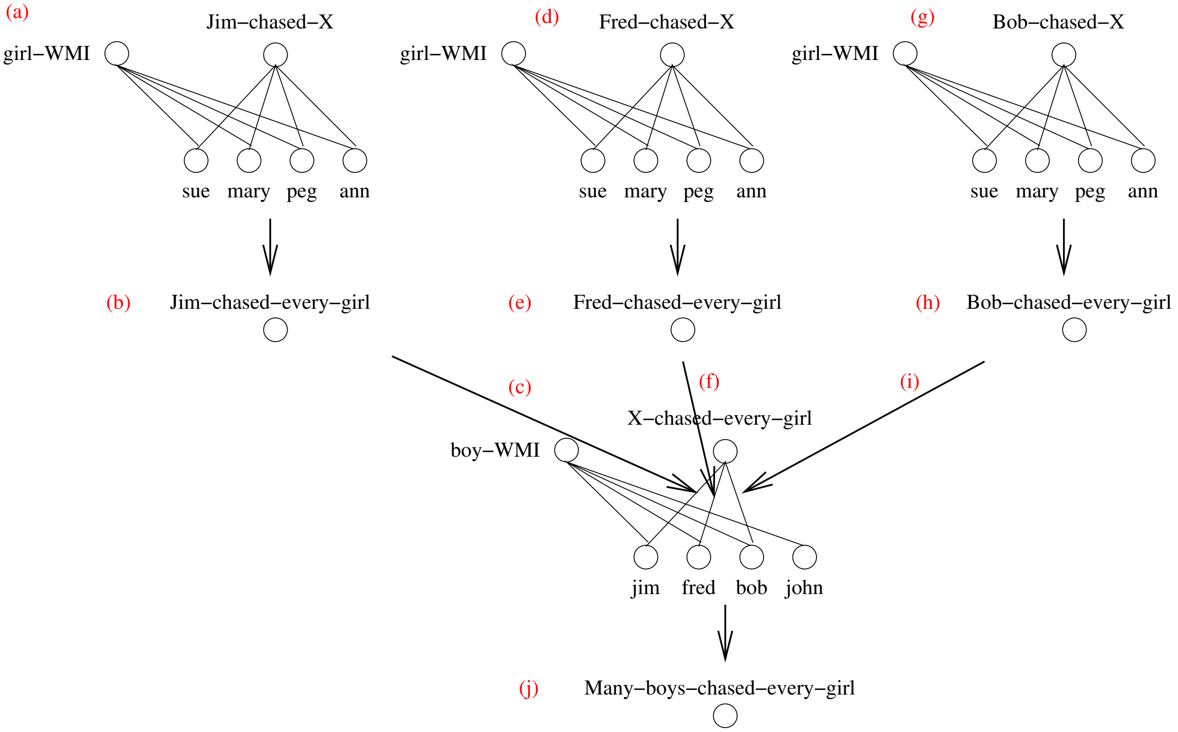


Figure 12.7: Semantic memory structures underlying *Many boys chased every girl* (with *many boys* outscoping *every girl*)

girls in the situation: he activates the WM individual 'girl', to establish the set of LTM individuals which are girls (in the current situation), and the collective property which emerges is the abstracted WM episode *Jim chased X*. Moreover, because he chased *every* girl, the numerosity of the set of activated girl LTM individuals does not reduce. This process is what underlies the generalisation 'Jim chased *every girl*', as already described in Section **??** (see Figure 12.7(b)). I now suggest that this WM episode can itself serve as the basis for a new entry in semantic memory, this time about *Jim*. We can abstract over the WM episode in the usual way, to create an abstracted episode 'X chased every girl'; then we can associate this abstracted episode with the LTM individual representing Jim (see Figure 12.7(c)). We might perform a similar operation for a number of boys, and associate the same abstracted episode 'X chased every girl' with other LTM individuals, say Fred and Bob (see Figure 12.7(d)–(i)). We can now express a generalisation over the boys in the situation (Figure 12.7(j)). First we evoke the WM individual 'boy', activating the set of all boys in the situation. Next we wait for a collective property of this set of boys to be selected. Say that the property which emerges is the abstracted WM episode 'X chased every girl', but that this property is only in fact possessed by some of the boys in the set. The numerosity of the set of boys will thus be reduced by a certain proportion. The generalisation over boys thus concerns a property which is *itself* a generalisation over girls. Of course, we could have created generalisations over boys to begin with, recording for each girl that 'many boys chased X', and then found that this property is one which all girls have. The basic idea is that multiply quantified propositions are constructed in several stages, each of which involves the encoding of a fact in semantic memory.

I assume that the most 'natural' multiply quantified propositions are those where the sequence of semantic-memory encodings is a natural one. For instance, say after watching John for some time the observer is drawn to note: 'my goodness: Jim chased every girl!', and then later is drawn to note something similar about several other boys individually. It is then quite natural for the observer to abstract further and note that many boys chased every girl. If the observer did not create the appropriate generalisations about individual boys, it might nonetheless be possible for him to deduce that many boys chased every girl, but this process would involve an effortful, sequential inspection of memory somewhat akin to an automated theorem-proving algorithm. People are notoriously bad at this (see e.g. Johnson-Laird, 1983 and much subsequent work).[7]

An important question is why it is that the linguistic expression 'Many boys chased every girl' fails to capture the relative order of the two generalisation operations. They are cognitively quite distinct. This is an issue which I [should be able to resolve by referring back to the chapter on the dp-clause interface].

---

[7] Better references are needed here.

12.9 Relative clauses and the semantic memory system

12.9.1 Distinguishing properties and meta-WM representations

When an episode is experienced, it is sometimes useful to represent its participants in a way which distinguishes them from other salient candidate individuals which might have participated but did not. For instance, if an agent has a choice of cup to grab, does the cup he grabs have any properties which distinguish it from the others? I suggest that we can choose to re-express a WM episode to make explicit any such distinguishing properties. For instance, say the cup which John grabs happens to be the only one which *he bought*. After experiencing the cup-grab episode, I suggest the observer can choose to look for a distinguishing property of the grabbed cup, and if one is found, form a more complex WM individual representation which makes reference to this property. This more elaborate WM individual representation does not just reflect the attentional processes required to establish the associated object. It also encodes an operation which occurs *after* the episode is experienced, which formulates a possible generalisation explaining the object's participation in the episode.

The operation which seeks for distinguishing properties of a participant in an episode can make use of semantic memory, which as we have seen, directly indexes individuals to properties. We might imagine, for instance, that all the candidate individuals, which might have participated in the event but did not, are configured to inhibit their associated sets of properties, while the participating object itself is configured to activate its set of properties. In this scenario, the property which is most active in the property competition layer will be that which most strongly distinguishes the participating object from the others. This could be an episode-level property or a simple perceptual property. I will not suggest a mechanism in detail; there are many ways it could be implemented.

One problem with evoking the distinguishing properties of an object participating in an episode is that it requires use of the WM episode medium which is already being used to represent the epsisode itself. In our example, for instance, we need a WM episode to represent 'X grabbed a cup', but we also need one to represent the distinguishing feature of the cup, namely that 'X bought the cup'. Representing both of these episodes simultaneously creates an ambiguity: it is impossible to distinguish 'X grabbed the cup which he bought' from 'X bought the cup which he grabbed', which encodes quite a different potential generalisation.

To overcome this problem, I suggest that an agent has working memory representations that can store *sequences of WM episodes* as well as sequences of sensorimotor states. Obviously these WM representations cannot be WM episodes themselves. I propose that there is a second, more abstract working memory medium in which sequences of cognitive operations can be stored (and rehearsed), where 'cognitive operations' include not only sensorimotor signals but also *transitory activations of WM epsiode representations.* Just as there is a medium for representing WM episodes and a medium for representing WM individuals, I will assume that there is a medium for representing **meta-WM episodes**, and one for representing **meta-WM individuals**. A meta-WM individual might hold a sequence of five operations: (i) attention to a location; (ii) establishment of a cardinality; (iii) activation of the 'cup' property complex; (iv) activation of the operation which evokes the most distinguishing properties of the cup; and (v) activation of the distinguishing property which is returned by this operation. The distinguishing property in our example is the abstracted WM epsiode 'John bought X', but it could also be a simple perceptual property, such as 'red' or 'big'. The meta-WM individual is illustrated in Figure 12.8. Basically it is a planned sequence of operations, just like a regular WM individual. The first three operations are just the same as for a regular WM individual; only the last two are different. The first of these is the operation of accessing semantic memory, and the second is the activation of a particular property—in this case the property is an abstracted (regular) WM episode (and is shown in red).

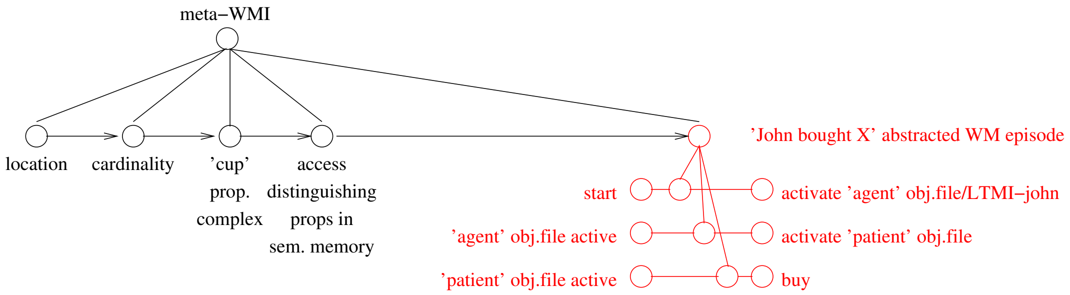


Figure 12.8: A meta-WM individual representing 'the cup which John bought'

I will not try to motivate meta-WM representations using paradigms in cognitive psychology or neuroscience. From these perspectives they are entirely speculative. Their eventual purpose, as might be guessed, is in an account of the semantic representations which underlie relative clauses.

Note that the evocation of a 'property' in the WM episode medium does not have any internal temporal structure: it happens in a single cognitive operation, because of the direct associations from LTM individuals to properties. However, because a property is a WM episode representation, it can still be *rehearsed* as a sequence, even if it is activated at a single point in time. The process of rehearsing a property may not have any 'semantic' significance—however, it may have an important role during the process of associating semantic representations with linguistic expressions. Individual properties are not associated with individual linguistic expressions, so the only way of mapping a property to linguistic expressions is to rehearse it as a sequence of sensorimotor signals, realised as a sequence of linguistic expressions, making use of mappings from individual sensorimotor signals to linguistic expressions. But note that this rehearsal involves another special mode of cognitive control. While we are rehearsing an evoked property, we are not rehearsing an experienced sequence of cognitive events, but rather using a device that allows us to express the content of the evoked property. This new mode of control will feature in the sensorimotor account of '*wh*-movement' which I give in Section 12.10.

12.9.2 'Bound variables' in quantified propositions

As discussed in Section 12.9.1, meta-WM individuals are used to redescribe a participant in a WM episode by referring to it by its distinguishing properties. Using meta-WM individuals, we can represent an episode like 'John grabbed a cup' as the (meta-)WM episode 'John grabbed [the cup that John bought]'. An episode of this kind can be the basis for a particularly interesting kind of quantified proposition. Say we observe several men grabbing cups, and in each case we note that the object grabbed is the cup which the agent bought. A generalisation can be made, to the effect that *Every man grabbed the cup which he bought*. The interesting feature of this generalisation is the pronominal reference *he*: rather than denoting some particular man in the domain of discourse, it denotes a different referent for each man which the sentence quantifies over. Linguists analyse a referring expression of this kind as referring to a 'bound variable' introduced by the quantifier *every*. A theory of quantification in natural language must allow for this form of reference, so I will introduce one final piece of machinery which supports it.

We have already seen how participants in a WM episode can be redescribed by meta-WM individuals, expressing their distinguishing properties. I now suggest that participants in a WM episode can also be redescribed as *re-activations of recently-attended objects*. In the case of 'John grabbed [the cup that John bought]', for instance, the reference to John inside the meta-WM individual happens to be a re-activation of a WM individual which is already active. We therefore do not need to describe the properties of this individual 'from scratch': we just need to indicate that one of the currently active WM individuals is being re-activated, and then specify which individual this is. (Recall from Section 10.6 that one of the key perceptual roles of a WM individual is to facilitate re-attention to an object which has already been attended to.) I suggest that a pronoun like *he* refers to an object indirectly in this way, by identifying a recently-active WM individual. (Pronouns specify a person, a number and a gender; I suggest that these are sufficient to match the location, cardinality and property complex which are related together by a WM individual.)

An example of a meta-WM individual containing this kind of indirect reference is shown in Figure 12.9 (the anaphoric reference to John is shown in blue). This meta-WM individual denotes 'the cup which *he* bought'. The referent for 'he' can be found anywhere in the WM episode which introduces the meta-WM individual. The whole (meta-)WM episode introducing this meta-WM individual now denotes something like 'John grabbed the cup which he bought'.

Now assume that this meta-WM episode can also be abstracted over, to create a property which can be linked to LTM individuals in semantic memory. In this case, we could link it to the LTM individual representing John. This will involve removing the reference to John in the episode—so instead of denoting 'John grabbed the cup which he bought', it denotes 'X grabbed the cup which he bought'. Importantly, note that the anaphoric reference does not need to be replaced: it is not a direct reference to John.

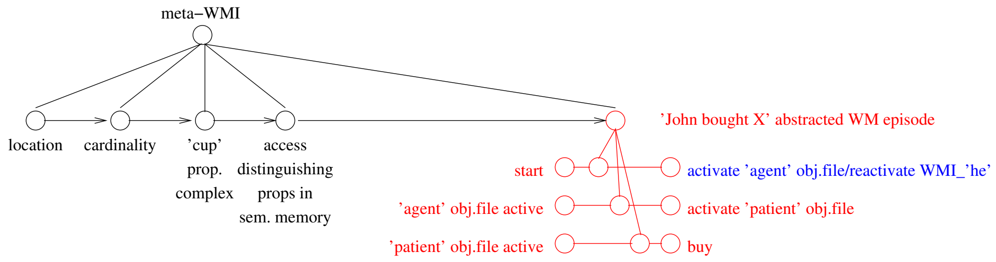


Figure 12.9: A meta-WM individual containing an anaphoric reference

Finally, suppose we observe several more episodes of cup-grabbing, and in each case, the agent grabs the cup he bought—and suppose that in each case we happen to create a meta-WM episode which redescribes the cup as the one which the agent bought, redescribing the agent anaphorically, as above. And suppose further that in each case, we then create an abstracted meta-WM episode, and link it to the LTM individual representing the agent. In each case, the abstracted meta-WM episode will be the same as above: 'X grabbed [the cup which he bought]'. This abstracted episode will then be linked to a number of different LTM individuals.

If the individuals associated with this abstracted episode also have some other property in common—for instance, say they are all men—then we are in a position to express another kind of quantified statement, of the form 'Q men grabbed the cup which they bought' (where Q can vary over 'all', 'every', 'most', 'many', etc). In each case, the quantified statement reflects a sequence of memory-access operations: firstly the posting of 'man' as a query to semantic memory, resulting in the establishment of a set of LTM individuals who are men, and secondly execution of the property-level IOR operation, resulting in activation of the abstracted meta-WM episode 'X grabbed [the cup which X bought]'. (Of course, this assumes that this abstracted episode is the most salient of the properties which are collectively associated with this group of individuals.) The number (or proportion) of LTM individuals which are have this property will then determine which quantifier is used in the statement.

12.10 *Wh*-movement in questions and relative clauses

[I may include a section on *wh*-movement here, which links the model of relative clauses just introduced to the model of questions given in Section 5.6.3.]

12.11 Summary

Chapter 13

Spatial cognition and the syntax of spatial PPs

13.1 Introduction

In this chapter, I consider the cognitive representations which underlie **spatial prepositional phrases** (**PPs**): phrases which describe spatial trajectories or spatial locations, typically in relation to a particular landmark. As usual, I will be focussing on concrete exemplars of spatial PPs, describing spatial representations which can be directly apprehended by the sensorimotor system. Some concrete spatial PPs are given below.

(13.1) On the table

(13.2) Onto the table

While the syntax of PPs is comparatively straightforward, their semantics is quite complicated. In fact, most of the current chapter is devoted to introducing a model of spatial cognition, which draws on a large body of research in psychology and neuroscience, and links back in several different ways to the cognitive models developed in earlier chapters. But at the end of the chapter I will also suggest how this model provides a framework for interpreting the syntax of PPs.

In the first part of the chapter, I outline an account of how 'places' are perceived and represented in both long-term and working memory. This account spans several sections. Section 13.2 considers the perceptual modalities which deliver representations of spatial location. Sections 13.3– 13.10 develop a model of long-term memory for spatial environments, and for the locations of objects within environments. Formulating this model touches on several other important issues, including the relationship between 'objects' and 'places', and an elaboration of the spatial aspects of reach-to-grasp actions.

In the second part of the chapter, comprising Section 13.11, I outline a model of locomotion actions, covering how these actions are planned, how they are executed, and how they are perceived.

In the final part of the chapter, I relate the model of spatial cognition to a model of syntax. In Section 13.12, I outline a model of the internal syntax of prepositional phrases. In Section 13.13, I argue that there is a natural way of interpreting the LF of a prepositional phrase as a description of a sequence of processes within the model of spatial cognition developed earlier in the chapter.

13.2 Spatial perception modalities

The brain has several specialised modalities for perceiving and representing space. These can be categorised along two dimensions. Firstly, different modalities use different frames of reference. For instance, some modalities use an allocentric frame of reference, while others use a frame of reference given in terms of the agent's eyes, head or motor system. Secondly, modalities represent place in two different ways. Some modalities derive a *holistic* representation of a region of space. For instance, there appears to be a specialised system for representing 'whole environments' in the parahippocampal place area. Other modalities represent a region of space as a *collection of places.* These can be thought of as maps. For instance, individual place cells in the hippocampus encode particular places within the current environment; together, these cells encode a map of the agent's environment.

13.2.1 Environment-centred space perception modalities

13.2.1.1 The parahippocampal place area

A key structure involved in our representation of place is the **parahippocampal place area** (or **PPA**) which was briefly introduced in Section 3.6.2.2. In that section I discussed the role of the PPA in representing individual 'spatial contexts' in the episodic memory system. However, it also plays a role in representing the agent's current environment. In this section, I will summarise what is known about the representations it computes.

The key finding is that the PPA responds to 'environments', not objects. In the experiments which show this effect, beginning with Epstein and Kanwisher (1998), environments are defined as having an 'extended spatial layout', of the kind that an agent can be 'in', and objects are defined as having a 'compact spatial layout', and as being entities which an agent can act *on.* Examples of environments are landscapes, empty rooms, or urban scenes containing buildings and streets; examples of objects are faces, and household/office implements. fMRI experiments show that the PPA responds to environments but not to objects; importantly, it responds just as strongly to scenes containing no objects (for instance an empty room) as to scenes containing objects. Interestingly, it responds to scene-like stimuli even if they have a spatial scale which would not permit an agent to navigate them. For instance, a scene consisting of an arrangement of objects on a desktop evokes a good response in the PPA (Epstein *et al.*, 2003), as does a scene consisting of a lego model of walls and corridoors (Epstein *et al.*, 1999).

The PPA's representation of an environment appears primarily to encode the spatial layout of the surfaces which objects can rest on within it, and of the boundaries which delimit it. Epstein refers to these surfaces as 'background elements' of the scene. If an observer is shown one scene twice in succession, the PPA responds less the second time, while if two different scenes with different spatial layouts are shown, the PPA's response to the second scene remains strong (Epstein *et al.*, 1999). The same paradigm also shows that the PPA responds differently to different *views* of a single scene (Epstein *et al.*, 2003). However, this viewpoint-specificity reduces as an observer gains experience about how a scene is spatially situated within a wider environment, at least in 'good navigators' (Epstein *et al.*, 2005.)

The PPA responds to both familiar and unfamiliar environments. However, it responds more strongly to familiar ones, which suggests that it has some role in the recognition of specific environments (Epstein *et al.*, 2007). However, Epstein *et al.* (2007) also found that another parahippocampal area called the **retrosplenial cortex** was *more* sensitive to the distinction between familiar and unfamiliar environments. They propose that both areas have a role in recognising specific environments. Their suggestion is that environments are identified in part by their spatial layout, but also by their spatial relationships to other known environments, with this latter component being represented in retrosplenial cortex. I will return to this idea in Section 13.4.3.3.

In line with the results just summarised, damage to the PPA or to the retrosplenial cortex results in impaired navigation abilities (Mendez and Cherrier, **?**). However, the PPA at least does not appear to be involved in *online control* of navigation, because it does respond any more to 'subjective-motion' environments than to static ones. (I will look at the perceptual routines involved in online control of motion in Section 13.11.)

Importantly for the model I will outline, the PPA and the retrosplenial cortex represent an *attended visual region.* The experiments described above present an environment to a viewer on a computer monitor; the environment to which the PPA responds only occupies a small portion of the observer's visual field. If the PPA responded to the observer's *actual* environment, in which the computer monitor is simply an object, we would not expect to see any changes in its activity as the monitor displays pictures of different scenes. The fact that we do see such changes means that input to the PPA can be gated by visual attention, in the same way that input to the object classifier can be.

13.2.1.2 The hippocampal cognitive map

I will now briefly review hippocampal representations of spatial location.

The first point to note is that hippocampal cells which are sensitive to location typically encode location in a coordinate system centred on the agent's local environment, i.e. in a manner which is insensitive to the location or orientation of the agent in relation to this environment. Interestingly, the definition of 'local environment' corresponds quite closely to that used to define the kinds of stimuli which the PPA responds to; an animal's environment is typically an enclosed area, defined by its *boundaries.* (Experiments investigating spatial representations in the hippocampus are frequently conducted on rats, in which case the boundaries tend to be the walls of a maze.) An influential study by O'Keefe and Burgess (1996) provided evidence that place cells represent the agent's location as a function of its distances to the various boundaries of its currrent environment. For instance, in a rectangular environment bounded by four walls, the location encoded by a given place cell could be defined as a simple function of four distances, in directions perpendicular to each wall.

The hippocampus can compute a representation of the agent's location which is stable over movements of the agent through the environment. I will call the function which computes this representation the **allocentric observer location function**. It is not yet clear how the function is computed. The hippocampus receives inputs from many perceptual modalities, but in primates, vision is of particular importance. Many models assume the existence of specialised visual pathways which create intermediary representations which provide input to the hippocampus. One influential model (Hartley *et al.*, 2000) proposes that the hippocampus receives inputs from a set of **boundary vector cells**, each of which is tuned to respond maximally when the animal is at a particular distance and allocentric orientation to a boundary. Boundary cells are theoretical constructs, hypothesised to underlie the sensitivity of place cells to the structure of an environment's boundaries. There has recently been some experimental evidence for such cells in the subiculum, an area which sits on a pathway from sensory inputs to the hippocampus (Barry *et al.*, 2006). Another likely source of input are **head direction cells**, found in the postsubiculum, which encode the orientation of the agent's head in relation to the environment (see Taube *et al.*, 1990). (I will refer to the function which computes this orientation as the **allocentric observer orientation function**.) Finally, it has recently been discovered that a parahippocampal region called the **entorhinal cortex** contains a population of **grid cells**, which evenly tile the agent's environment with hexagonally packed grids at various spatial frequencies (Fyhn *et al.*, 2004). These cells also provide input to the hippocampus. They are likely to be involved in updating the agent's location using dead reckoning; they may also provide more complex spatial information, supporting some form of coordinate system. A useful account of how these inputs could be combined is given by Jeffery, 2007).

How does the hippocampus learn to map these inputs to representations which are stable over the agent's movements and changes in orientation? A very interesting recent suggestion by Franzius *et al.* (2007) makes use of the constraint that the agent's allocentric location can only change relatively slowly from moment to moment, even though the agent's raw retinal input can change quite dramatically as he changes orientation within his environment. Franzius *et al.* show that imposing this constraint on the function which processes sensory input is in fact sufficient to oblige it to compute an allocentric representation of the agent's location. They simulate an animal moving through a rich visual environment and altering its head direction, and learning a function whose inputs are a collection of boundary vectors. If the animal moves relatively slowly and makes relatively large alterations to its head direction, the function learns an allocentric representation of the animal's location in the environment, of the kind found in place cells. If the animal moves relatively quickly and makes minimal alterations to its head direction, the function learns an allocentric representation of head direction, of the kind which is found in the postsubiculum, as described above. The constraint of a slowly changing representation allows both the allocentric observer location function and the allocentric observer orientation function to be learned.

Some caveats must now be given. The above account assumes that the hippocampus provides a map of agent locations, in which cells encode positions which can be reliably inferred from the geometry of the environment's boundaries. However, the picture is much more complicated. For one thing, as already mentioned in Section 3.6.2.1, hippocampal cells in primates and humans can represent the location which the agent is *attending to.* They can also represent particular combinations of objects and locations; in other words, they are involved in representing memories of the spatial locations of objects. (See Rolls *et al.*, 1997a; Rolls, 1999.) For another thing, the location a hippocampal cell encodes cannot always be predicted from the geometry of the environment's boundaries. As already described in Section 3.6.2.2, when the agent moves from one environment to another, many hippocampal cells are **remapped** to new locations; this happens even if the shape of the new environment is the same as the old one. These two findings are likely to be related. If hippocampal cells are involved in memory for object locations, there must be a way of individuating environments even if they have the same shape, because they may have different contents at different times. In Section 3.6.2.2 I discussed the proposal that the hippocampus also receives input from a 'context' representation, so that connections between objects and locations can be made selectively in different environments. It would make sense if the holistic representations of environments described above, in the PPA and retrosplenial cortex, provided this input. I have not seen this proposal explicitly, but I will assume it henceforth. Note that remappings do not appear to affect an animal's navigational abilities (see e.g. Jeffery *et al.*, 2003). It thus seems that the representations of object location memory can be dissociated from those which represent the agent's current location. Finally, it may be that some hippocampal representations encode information about an agent's locomotion goals. I will defer discussion of this data until Section 13.11.3.2.

13.2.2 Object- and agent-centred space perception modalities

It is also useful to revisit the account of the spatial representations which subserve perception and execution of reach-to-grasp actions, as described in Chapter 2. These representations can also be classed as holistic or map-based.

13.2.2.1 Holistic spatial representations of objects and agents

The object categorisation system provides an obvious holistic spatial representation of objects. The representation it provides is 'spatial' in the sense that objects are categorised in part by their shape. We recognise objects in large part because of the highly derived shape representations delivered by this system.

We also have special modalities for computing holistic spatial representations of *agents.* We can use proprioceptive information to generate a representation of the configuration of our own body, and internal models and efferent copy to generate a representation of how this configuration is changing (see Section 2.5.2). To represent the body of an observed agent, we use a specialised biological motion recognition system (see Section 2.7.2.1), which takes input from a combination of form and motion cues.

13.2.2.2 Motor maps for representing objects and agents

Now recall from Section 2.5.2 that the parietal 'reach pathway' maps visual and somatic inputs onto a representation of the location of potential targets in the agent's peripersonal space, in a coordinate system defined in terms of motor commands to the arm. This pathway can be thought of as computing a motor map of the agent's perispace—or more precisely, as computing one motor map for each of the agent's *motor systems*. A motor system might be a single limb, but it could also be a pair or set of limbs working in coordination. Each motor system which can be 'deployed' to points in the agent's perispace will be associated with a map of points in the perispace, represented as motor commands (or to use the terminology of Section 2.5.2, as **movement vectors**).

Recall also that the parietal 'grasp pathway' takes a single attended object and represents it as a set of grasp affordances, i.e. as a set of goal hand motor states—see Section 2.5.3. Again, we can think of these affordances as specifying a map of 'places' within the target object—or more precisely as a set of **opposition axes** within the object.

In summary, there are holistic representations of both agents and objects, and there are also representations which encode agents and objects as maps of places.

13.3 The relationship between environments and objects

A useful starting point in developing a model of spatial representations is a model of the relationship between environments and objects. We have already proposed a means by which the locations of objects within an environment can be represented. But note that objects have spatial structure, and if they are considered at the right *spatial scale*, they can be thought of as environments. For instance, consider a small coffee table. It is an object as defined in Chapter 2, in the sense that we can attend to it, categorise it and pick it up. However, if we were the size of Tom Thumb, it would be an environment which we could walk across, with boundaries (the edges of the tabletop) and possibly obstacles (any large objects on the table which are hard or impossible for Tom Thumb to move). Likewise, what we think of most naturally as an environment can be an object when considered from a different spatial scale. Consider a garden. We categorise this as an environment, rather than an object. We probably use a specialised perceptual mechanism for recognising 'environments' like gardens when we are in them, as will be described in Section 13.10.2. But for a giant, a garden might be something which is typically seen from a certain distance, and which the giant recognises by its typical shape. (The giant might even be able to pick the garden up, so it can also be represented for him in terms of its reach-to-grasp affordances.) In fact, many things in our world naturally switch between being classified as objects (when we are far from them) and as environments (when we are in them). Houses and cars are two good examples. In summary, the distinction between objects and environments is to do with the spatial scale which is adopted when representing them, rather than with intrinsic properties of things in the world. This means that the key to developing a model of how environments can be nested is to understand what it means to 'adopt a certain spatial scale' when spatially representing the world.

In the current section, I will develop a model of the relationship between objects and environments, which is grounded in the idea of spatial scale as used in the account of visual object perception given earlier in this chapter. The key idea is introduced in Section 13.3.1. Later sections will consider different ramifications of the idea.

13.3.1 A perceptual model of the relationship between objects and environments

[I need to add several figures in this section, to make the ideas clearer.]

We have already made use of the idea that different spatial scales can be adopted when attending to and categorising stimuli in the world. Recall from Section 10.8.3.1 that there are several distinct saliency maps, operating at different spatial frequencies, and that at any time, one of these maps 'controls' at what spatial frequency objects are classified. This model of hierarchy in visual attention can be extended to provide an account of the distinction between objects and environments.

As already discussed, the brain has a specialised perceptual modality for apprehending 'spatial environments', which is distinct from its modality for categorising objects: the environnment perception modality introduced in Section 13.2.1.1, and localised in the PPA. As discussed in that section, the environment perception modality can be deployed selectively to salient retinal regions, and at different spatial frequencies—in other words, it is amenable to exactly the same kind of gating as the modality which classifies objects. When the agent has established a salient retinal region, should he deploy the environment perception modality, or the object classification modality? Or can he perhaps deploy both? I will suggest some answers in this section.

13.3.1.1 Objects and environments: competition within spatial frequencies

To begin with, I propose that within a given spatial frequency, the object and environment systems compete with one another, so a selected region cannot be classified *simultaneously* as an object and as an environment. Competition is driven by several factors. One of them is size. If the portion of the world picked out by the salient region is big enough for the whole agent to navigate through, this encourages it to be classified as an environment. Another is distance. If the portion of the world corresponding to the salient region is navigable by the agent and also close at hand, it is even more likely to be classified as an environment. Trading off against these cues is shape. Certain objects have characteristic shapes; if the portion of the world corresponding to the salient region is far enough away for its shape to be apparent, this encourages it to be classified as an object. Animacy also biases classification towards objects—objects often move, environments tend not to.

I assume that there are some cases where these criteria are well balanced, in which it is fairly arbitrary whether a region is classified as an object or as an environment.

13.3.1.2 Objects and environments: a relationship between adjacent spatial frequencies

What is not arbitrary is the relationship *between* objects and environments. For instance, if a region has been classified as an object, the environment *of that object* must be a portion of the world which contains it. I propose that there is a systematic relationship between the region and spatial frequency at which the environment perception modality operates and those at which the object categorisation modality operates. Recall from Section 10.8.3.1 that the controlling saliency map corresponds to a single region in the 'super saliency map' at the next spatial frequency down. I suggest that the environment perception modality is associated with this same region, and with a commensurate spatial frequency. If the controlling saliency map covers the full visual field, then the agent is 'looking around his environment', and his environment perception modality delivers a representation of the spatial structure of this environment. But now assume the agent selects one region in his visual field to attend to. As just discussed in Section 13.3.1.1, this may be classified as an environment—for instance, some distant cliffs, or a garden glimpsed through a window—or it could be classed as an object. In either case, the agent is free to establish a sub-saliency map within this new region. If the region was classed as an environment, this operation can be carried out directly, because points in the saliency map will be associated with spatial locations in the environment. But if the region was classed as an object, *it must be re-established as an environment* before we can 'look within it' in this way. It is only then that we have a representation of the spatial structure of the object which allows us to characterise 'locations' within it.

In summary: I suggest that our concept of the relationship between objects and environments is grounded in an account of spatial scales in the perceptual system. The key idea is the idea of sub-saliency maps introduced in Section **??**: if an object is associated with a region in a saliency map at a certain spatial scale, then the environment of that object is associated with the whole saliency map. In the remainder of this section, I will discuss some ramifications of this idea.

13.3.2 Attentional exploration and attentional capture

If I arrive in a new environment, I might spend some time looking around, to see what it contains. This kind of exploration presumably involves an orderly traversal of the environment, driven by inhibition-of-return. In this situation, I establish the environment first, and then consider what is in it.

However, consider what happens when my attention is *captured* by an object—perhaps by a sudden movement or noise. I must presumably attend to and classify the new object. (Its animacy makes it likely to be classified *as* an object.) But I may also have to shift my attention to a new *environment* at the same time. For instance, say I am looking around my office when my attention is drawn to a dog outside in the garden, visible through the window. In this situation, I must establish the dog as an object, but also the garden as an environment.

I assume that in a case of attentional capture, there are two 'winning' retinal regions: one associated with the object capturing attention, and the other associated with the same region, but at a lower spatial frequency. This latter region will function as the environment within which the object is localised. Object and region can be established in parallel, but the spatial relationship between them (*in*, *on* and so on) can only be determined when they are both established.

13.3.3 Classification of cluttered environments

An environment is often cluttered with objects. How is it recognised in the presence of these objects? The large spatial features of the environment are likely to be resilient to clutter. However, many planar environments are largely characterised by their visual texture. For instance, we are likely to recognise lawns, beaches or bodies of water, by characteristic textures. The texture of objects *in* these environments might therefore get in the way.

One simple scheme emerges out of the hierarchical model of salience just outlined. We already know that the environment classification system only receives input from a particular salient retinal region. If we also stipulate that input to the environment classification system is blocked at all points in the sub-saliency map associated with the region it is classifying, we have an effective means of ignoring the objects within the region.

13.3.4 Figure/ground reversals

The relationship between an object and its environment can also be referred to as the relationship between a **figure** and its **ground**. These latter terms are more general, as they also apply to drawn stimuli, which consist simply of coloured lines and regions. There is a well known phenomenon called **figure-ground reversal**, which can be modelled using the current conception of objects and environments. The phenomenon is illustrated in Figure 13.1. Does the figure show a cup on a black background, or two faces on a white background? Cues like distance are not available, and the remaining form cues

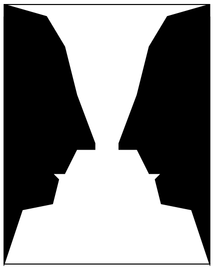


Figure 13.1: Figure-ground reversal

are ambiguous. It is not possible to perceive both figures simultaneously—the observer's percept alternates between the cup and the faces.

In the model currently being proposed, this effect can be attributed to the constraints which require an object (figure) to be established in the context of its environment (ground). In each case, the environment is the rectangle enclosing the picture, which functions as a saliency map; the object is something 'within' this environment, i.e. a salient region within this map, established at a higher spatial frequency. There are three regions in the saliency map for the image, defined by the two texture boundaries.[1] Attention is allocated to the different regions, and object categories are assigned to the attended regions. Importantly, when an object is established, its environment must also be established. If the object established is the cup, its environment is constrained to be the remainder of the image, i.e. the black region. Note the environment classification system does not receive input from the figure region, so only operates on the black visual texture. Conversely, if the object established is one of the faces, its environment is the white region. Again the environment classification system does not receive input from the figure region, and so only operates on the white texture. The model thus gives an account of why it is impossible to establish the cup and the faces simultaneously.

13.3.5 Objects as environments: the haptic interface revisited

Consider a situation where an agent attends to a cup, and then creates a sub-saliency map associated with the cup, in order to 'look within' it. In Section 13.3.1.2 I suggested that the agent must at this point re-establish the cup 'as an environment', within which the objects found in the sub-saliency map can be localised. Note that this environment will be too small for the agent himself to navigate through. In what sense, then, is it an environment? How is it spatially defined?

I suggest that an object can be thought of as an environment navigable by the agent's **effectors**—for instance, by the agent's hands or fingers. Effectors can be thought of as entities that undergo locomotion actions which are similar in many ways to those which the whole agent undergoes. (I will refer to a 'complete' agent as an **autonomous agent**, in contrast to **effector agents** such as hands and fingers.) The agent's reach actions cause his hands to travel through his peripersonal space, avoiding obstacles and arriving at targets. When the agent's hand touches a target object, his hands and fingers can travel around the object, guided by his sense of touch. These actions are routines defined in the 'haptic modality' introduced in Section 2.5.4. When touching an object, hands and fingers move through an environment of surfaces and obstacles, just like an autonomous agent. However, the means by which effectors can 'travel' through their environments are subject to a different set of constraints than those for an autonomous agent. Some constraints are relaxed. For instance, hands and fingers are not constrained by gravity to rest on a supporting surface: they can fly around the environments they interact with, moving above

---

[1] I should refer to Hayden's implementation here. There's some sort of effect which means that the ambiguity is only found when 'figures' are large and quite close, I think.

or below them, and they can assume a wide range of stable rotational positions. Moreover, hands and fingers can interact with their environments by applying *opposing forces*: fingers can grip an object, or a pair of hands can squeeze an object. This is not something an autonomous agent can do. There is also a way in which the movement of effectors is more constrained than that of autonomous agents. An effector is connected to a superordinate motor system—a hand, or an arm, or a whole body. If these appendages do not fit into the environment which the effector is exploring, this will limit its range. The appendages also place constraints on the ease with which the effector can rotate about its own axis. An autonomous agent can turn a full circle with ease; an effector cannot.

I suggest that the spatial characterisation of an object as an environment is given by the types of action it affords to motor subsystems of the appropriate scale. I will discuss this idea in detail in Section 13.7.2.

13.4 A preliminary model of spatial LTM

Having introduced the perceptual modalities involved in generating spatial representations, and sketched an account of the relationship between objects and environments, I will now outline a preliminary model of spatial long-term memory—in other words, of how the spatial relationships between different individual objects and environments are represented and stored. The model will be revised and extended in Sections 13.5–13.11.

13.4.1 LTM environments

To begin with, I will assume that the agent maintains LTM representations of the individual environments he is familiar with. We have already introduced the idea of 'individual spatiotemporal contexts', in the discussion of episodic memory in Section 3.6.2.2. However, for the current discussion, we will focus on the spatial components of these representations. I will call individual spatial context representations **LTM environments**.

I have already mentioned in Section 13.2.1.1 that PPA and retrosplenial cortex have a role in the recognition of individual environmnents. I assume that LTM environments are stored in these areas.

13.4.2 Representing the location of individual objects within an environment

We have already introduced a model of how individual objects are represented in long-term memory. Recall from Section 3.6.4.1 that an **LTM individual** is an assembly in perirhinal cortex which represents an individual object (or agent) familiar to the observer. I also suggested in Section 3.6.4.2 that long-term memory for object locations is implemented by patterns of hippocampal activity whose significance is given by the representation of 'spatiotemporal context' currently in force. In this section, I will provide a more detailed model of memory for object locations.

13.4.2.1 The external object location function

To begin with, recall from Section 13.2.1.2 that the hippocampus contains cells which respond when a particular location is *viewed*, regardless of the location of the observer. I will call the function which generates these representations of attended locations the **viewed location function**, or if the attended location contains an object, the **external object location function**. This function differs from the allocentric observer location and allocentric observer orientation functions, which generate representations of the observer's *own* location and orientation in the environment. However, these latter functions combined allow a vector to be specified on which the observed object or location must lie. All that is needed to specify a point on this vector is an egocentric measure of the distance from the observer to the attended location. In other words, the external object location function can be generated from the same boundary vector inputs as the observer location and orientation functions, supplemented with a measure of distance-to-attended-location. Franzius *et al.* (2007) show that their constraint on slowly changing location representations allows an external object location function to be learned from these inputs, under conditions where the observer allocates sustained attention to stationary objects while moving through the environment itself—again, a very neat result.

Note that an allocentric representation of the observer's direction of attention cannot be read simply off the orientation of the observer's body in the environment. The angle of the observer's eyes in relation to his body must also be considered. In addition, we must recall that the observer maintains a *saliency map* of locations to attend to. As discussed in Section 2.4.4, there seem to be several copies of this map, some centred on the retina, others on the head. The external object location function must take input from the observer's allocentric orientation in the environment, but also from the angle of the agent's gaze in relation to his body, and from the horizontal location of the associated region in the saliency map.[2] In summary, the external object location function maps points in the observer's saliency map onto points in the observer's current environment.[3]

One interesting question concerns whether the mapping from the saliency map to an environment-centred map is effected in parallel for all points in the map, or for one point at a time. There is no direct data bearing on this question, as far as I know. However, there is good evidence that the mappings between retinotopic, head-centred and body-centred saliency maps are implemented in parallel in the parietal cortex (see Sections 2.4 and 2.5.2), at least for a small number of salient locations. Indeed, there are some models of the neural implementation of coordinate system transformations which assume that locations are represented in several coordinate systems simultaneously (see especially Pouget and Sejnowski, 1997). Perhaps the main difficulty for a parallel implementation of the external

---

[2] Note also that the *vertical* angle of the currently selected region in the saliency map in relation to the horizontal plane can also function as a cue indicating the distance of the location or object which projects to this point.

[3] Somewhere in here I should cite Spratling's (2009) model of how retinotopic maps of locations are converted to head-centred and then body-centred coordinate systems. This model is attractive, in that a whole set of salient retinal locations are mapped *in parallel* to points in these more stable coordinate systems.

object location function is how to compute the distance of each point in the saliency map in parallel. Some representations of distance are computed in parallel—the binocular disparity information used in stereopsis is a good example (see e.g. Livingstone and Hubel, 1988). But others can only be computed point-by-point—for instance, knowing the category of an object allows us to use information about its normal size to help estimate its distance, but objects can only be categorised one at a time (see Section 2.4). In conclusion, it seems likely that there is a parallel implementation of the mapping from the retinal saliency map to an environmnent-centred map, and also a serial one, which provides a more accurate estimate.

13.4.2.2 The orienting function

The external object location function maps the agent's current allocentric location and orientation onto the point in the environment which he is attending to. The same learned correspondences which enable this function also enable the agent to attend to any point in the environment. I assume another function, called the **orienting function**, which takes the agent's current allocentric location and orientation, plus an arbitrary point in the environment, and returns a **goal allocentric orientation**. The agent's current and goal allocentric orientations together provide input to a motor controller which generates an **orienting action**, which might be defined as the activation of a point in the saliency map, or as a saccade, or as a rotation of the head or the body, or as a rotation of the whole agent, or possibly as a combination of several of these operations. The motor controller which generates such movements is likely to be implemented in parietal cortex. We have already discussed the mechanism which generates top-down activation in the saliency map; see Section 2.4.2. The parietal cortex also contains cells with gain fields reflecting the agent's orientation in his environment (see e.g. Snyder *et al.*, 1998 for evidence in macaques) and which are involved in controlling orienting movements of the head (see e.g. Klam and Graf, 2006). In summary, I will assume that the orienting function takes allocentric representations of the agent's current location and orientation from the hippocampus and parahippocampal regions, plus an allocentric representation of another arbitrary region in the current environment from the hippocampus, and generates an orienting action which establishes the retinal location onto which this region projects as the new most-active point in the saliency map.

13.4.2.3 Object location memory

As suggested in Section 3.6.4.2, I propose that the location of an object is encoded as an association between an LTM individual and a hippocampal place representation. Such an association is created when the observer looks at an external object. In this situation, the object recognition function delivers an LTM individual, and the external object location function delivers an environment-centred place representation. Hippocampal 'view cells', which encode combinations of object identity and environment-centred location (see Rolls *et al.*, 1997a; Rolls, 1999) are presumably involved in implementing these associations.

Recall from Section 3.6.4.2 that such associations must be completely dependent on context representations. A given hippocampal cell may be used to represent locations in many different environments. If the agent associates a particular object with this location in one environment E1, there will be an association between the hippocampal cell and a particular LTM individual. However, when the agent moves to another environment E2, this association should no longer hold. Similarly, memories for object location should be dependent on temporal context representations. An object might be in one location at time T1, but be somewhere else at time T2. Associations between LTM individuals and hippocampal place units must thus be gated by specific context representations. I assume that 'the current spatial context' is represented by an active *LTM environment*, as introduced in Section 13.4.1. I will consider temporal contexts in more detail in Section 13.9. In summary, we can envisage an **object location memory function**, which takes a context representation (comprising an LTM environment and a temporal context) and delivers a set of associations between hippocampal place units and LTM individuals. Note that when a particular association is in place, activating an LTM individual will generate an orienting action which directs the agent's attention to the associated location. A network which implements this function is sketched in Figure 13.2.

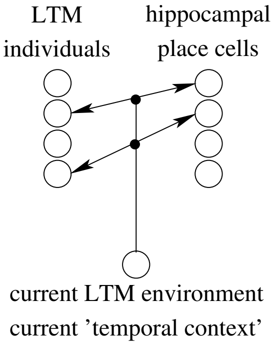


Figure 13.2: A simple model of object location memory

13.4.3 Representing the spatial relationships between individual environments

In Section 13.4.1 it was proposed that specific environments are represented as individuals in LTM. However, it says nothing about how the spatial relationships *between environments* are represented in LTM. These spatial relationships can be of different kinds. The most important relationship between environments is **accessibility**: can an agent get (directly) from one environment to another other? And if so, how? I will distinguish two types of accessibility, and propose a means for representing each in LTM.

13.4.3.1 Representation of adjacent environments

Firstly, two environments can be spatially adjacent. The simplest situation to consider here is one where each environment is reasonably spatially enclosed, and where there is a reasonably localised point at which they connect. For instance, we might consider two rooms, linked by a doorway. For an agent who knows about the layout of these two rooms, each room will be represented in LTM by an LTM environment. Place cells firing in the agent's cognitive map have a representation which is to be understood in relation to the currently active LTM environment. I will call locations within a given environment **places**. (The agent's *representations* of 'places in the current environment' are therefore given by place cells.)

How does the agent represent the layout of the two rooms? I propose that alongside the object location memory function introduced in Section 3.6.4.2 there is an **environment location function** which given an active LTM environment, delivers a set of associations between cognitive map places and *other LTM environments*. The associations for any given environment give the places in that environment at which there are adjoining environments. If the agent moves to one of these places, he will be able to enter a new location.

Note that to access a new environment involves adopting an appropriate *orientation* as well as an appropriate place. There is good evidence that agents maintain an environment-centred model of their current orientation as well as of their current place. In rats, this appears to be maintained in parahippocampal area called the **postsubiculum** (Taube *et al.*, 1990). In our two-rooms example, imagine that the agent is in Room 1, and reaches the doorway. Entering Room 2 requires adopting a particular orientation: if the agent has his back to the doorway, the adjacent environment will not be established. Once the agent establishes the right orientation, his perception will be of Room 2; essentially he has 'entered' the second room by adopting a suitable position and orientation within Room 1. We must therefore extend our environment location function a little. I will define the agent's **current spatial state** as a tuple of a particular place and a particular orientation. The environment location function now maps the current LTM environment onto a set of associations between *spatial states* and new LTM environments.

It appears that a rat's representation of head direction is remapped when entering an adjacent environment, just as its representations of places are. So in order to represent information about the relative *configuration* between two environments, it is probably important to extend the environment location function a little further, so that when a new LTM environment is activated, the rat also activates an appropriate new spatial state (including place and orientation).

A network for representing neighbouring environments is shown in Figure 13.3. The network contains a link from spatial state S1 to spatial state S2, and a link from spatial state S1 to environment E2 (both shown in red). Both these links are gated by environment E1, and so are only active when E1 is the current environment. If the agent is in environment E1 and arrives in state S1 (a particular place and orientation), this will cause his current environment representation to be *updated* to E2, and his current state representation to be updated to S2 (i.e. his current place and orientation representations will also be updated

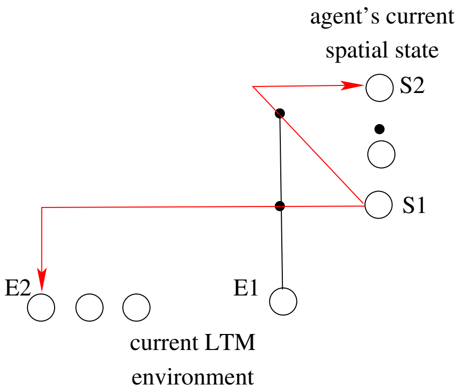


Figure 13.3: A model for encoding the location of neighbouring environments

to conform to the new environment).

13.4.3.2 Representation of hierarchically nested environments

As described in Section 13.3.1, an object within one environment can be attended to as an environment in its own right. This means that one environment can contain other environments. How are such hierarchical relationships represented in spatial LTM? As a general case, let us assume that there are two environments, E1 and E2, such that E2 is 'in' E1. Assume further that E1 and E2 are each represented by a unique LTM environment. What is the relationship between these two LTM environments?

First, as for adjoining environments, we need to specify the *place* in E1 where E2 is—i.e. where one has to go to in E1 if one wants to get into E2. Note that the 'size' of this place will depend on the size of E2 in relation to E1. As it happens, hippocampal place cells are defined at a range of spatial frequencies, just like cells in early visual areas (see e.g. Jung *et al.*, 1994; Kjelstrup *et al.*, 2007). So representing location consists in picking a place cell at (a) the right location, and (b) the right spatial frequency.

Second, there might also have to be a special *spatial scale* which you adopt in order to transition from E1 to E2. As an agent, 'adopting a smaller spatial scale' might involve shifting to a motor control regime in which one's hand is monitored as an independent moving entity. As an observer, it involves establishing a representation which assumes an agent of a certain size. I assume that the appropriate spatial scale can be 'read off' the spatial frequency of the place cell which represents E2's location. (It should certainly be correlated with this, at very least.)

Third, there might be arbitrary motor actions you have to perform to get into the environment. For instance, to get onto a table requires a 'climbing' or 'scrambling' action; to get onto a low platform requires a 'step up'. These are locomotion actions, in the sense that they move the agent, but they are primarily defined in terms of the agent's *configuration.* The early stages of such actions are quite likely to be defined on the new environment *established as a reach target.* For instance, to climb onto a table requires the agent to bring his limbs into contact with the table in certain ways. Their latter stages involve establishment of a certain configuration of the whole agent in his new environment—for instance, the configuration of standing or kneeling on the table. These special environment-changing motor actions can thus be thought of as operations which bring two frames of reference into correspondence, just as regular reach-to-grasp actions do. I will call such actions **agent reconfiguration actions**. These actions should also feature in the definition of the relationship between one environment and a nested one.

Sometimes an embedded environment is primarily defined by this reconfiguration action. For instance, a chair is an environment we can 'get into' by a particular whole-body reconfiguration, but once 'in' a chair, there is very little scope for locomotion.[4]

In summary: at a first approximation, the function which delivers the environments nested in the currently active LTM environment should associate a new LTM environment with a tuple comprising (a) a place in the current environment, (b) a relative spatial scale, and—optionally—(c) an agent reconfiguration action.

It is also important to consider how the embedding environment E1 is indexed to the embedded environment E2. If the observer is simply *attending* to the embedded environment, then re-establishing the wider environment just involves choosing a larger spatial scale for the scene perception modality. However, if the observer is *in* the embedded environment, then something more is required. Basically, the embedding environment will be associated with *the boundaries* of the current environment, so what is required is to navigate beyond these boundaries. An agent reconfiguration action might be required here too, such as 'climbing down' from a table or 'stepping off' a platform. I will assume that there is a function which maps the embedded LTM environment to a set of associations between the embedding environment and a tuple of (a) a boundary representation and—possibly—an agent reconfiguration action.[5] Note that this function means that an agent can always 'pull back' to a wider embedding context by establishing the LTM environment associated with the boundaries of his current LTM environment.

13.4.3.3 A possible role for the retrosplenial cortex in storing spatial relationships between connected environments

Where in the brain are the spatial relations between ajoining or nested environments stored? There is some evidence that the retrosplenial cortex has a role in holding this information. Recall from Section 13.2.1.1 that Epstein *et al.* (2007) identified two parahippocampal areas, the parahippocampal place area and the retrosplenial cortex, as being

---

[4] Other than a little shuffling, I suppose.

[5] One issue I am not considering is *whereabouts* in the embedding environment the agent arrives when emerging from the embedded environment. If the embedded environment has spatial extent, then different exit points will leave the agent at different places in the outer environment. If the embedded environment is closed, and the agent has been in it too long for path integration to be of any help, it may simply be that the agent has to re-establish this perceptually. If the embedded environment is open, it may be that the agent has enough perceptual information about the boundaries of the embedding environment to remain oriented to it, perhaps by periodically switching between environments.

preferentially activated by familiar environments, and thus as being involved in representing individual environments. However, Epstein *et al.* also found that the familiarity effect was considerably stronger for retrosplenial cortex than for PPA. They propose a model in which environments are individuated in two ways: firstly by a characteristic spatial layout, and secondly by characteristic connections to other known environments. They suggest that PPA is primarily involved in individuating objects by the first means, while retrosplenial cortex is also involved in individuating objects via the second means. Thus a good case can be made that knowledge of the spatial links between individual environments is stored in retrosplenial cortex.

There is another account of the role of the retrosplenial cortex, which proposes that it is primarily involved in transforming allocentric spatial representations in the parietal cortex into environment-centred hippocampal representations (see e.g. Burgess *et al.*, 2001). Some of the evidence used to support this model could equally well support a model in which retrosplenial cortex stores the spatial relationships between LTM individuals and places. For instance, Aguirre and D'Esposito's (1999) finding that lesions of retrosplenial cortex lead to difficulties in forming or recalling links between landmark identity and directional information are equally well explained. However, the strong projections from retrosplenial to posterior parietal cortex must also be explained. One possibility is that these are involved in representing the reconfiguration actions which can be needed in order to move from one environment to another. Such actions often involve reaches, or arbitrary limb movements. It is plausible to envisage that representing reconfiguration actions will require links to the parietal cortex. Note that the retrosplenial cortex also projects to the prefrontal cortex (see e.g. Shibata *et al.*, 2004 for data from rats), which makes sense if reconfiguration actions are defined as action sequence plans, as proposed in Section 3.2. However, this idea is very speculative.

13.4.4 Special environments: trajectories, sub-environments and configurations

In this section I will introduce three special types of environment, which are less easy to conceptualise as objects, even at a coarser spatial scale. I will begin by considering *paths*, which have some interesting special characteristics. Then I will consider components of environments which are referred to using the environment's own coordinate system, such as the top of a table, or the corner of a room. Finally I will consider 'configurations of objects' as environments. Each of these environments is a special case, which must be related to its containing environment by special methods.

13.4.4.1 Paths as environments

A field might have a path running through it. We can talk about an agent being *on* the path, or moving *onto* the path; by these criteria, it seems like an environment nested within the field environment. However, it is hard to think of the path as a normal object within the field, like a tree or a car. In fact, a path might only be defined as an absence of obstacles, rather than any particular form or texture. While it makes sense to talk about the path being 'in' the field, it does not appear to be nested *as an object* within the field. How should a path be characterised as an environment?

I suggest it makes most sense to think of the path as a *trajectory* in the environment: it defines a route which an agent could take. We have seen in Section 3.7.1 that sequences of place cells can be stored and recalled; in Section 13.11.3.1 I will expand on this idea in the context of spatial navigation, and outline evidence that trajectories are a natural unit of information storage for the hippocampus. In the present context, I propose that there is a special class of sub-environments whose location within their nesting environment is given by a sequence of place cells. I will call these environments **path environments**.

An agent can navigate along a path environment. However, a path environment can also be established visually. Observers can execute **visual routines** (Ullman, 1984) which track elongated visual stimuli such as lines or paths. An observer can represent the trajectory of a moving object in an environmnent-centred frame of reference, as will be discussed in Section **??**. I suggest that these are the perceptual modalities by which the location of path environments are established.

Note that many 'long thin' objects can be usefully characterised as trajectories. For instance, a fence is an object which can presumably be recognised by its distinctive form. However, the location of this object is best defined as a trajectory.

13.4.4.2 Sub-environments

We can talk about someone being in the *corner* of a room, or of a cup being on the *edge* of a table. By this criterion, corners and edges are environments in their own right. Prima facie, we might think that a corner stands in the same relation to the containing room environment as other objects in the room. But again, a corner is not an object *in* the room; it is a *part of* the room. It cannot be moved to different locations within the room, like ordinary objects. It cannot be referred to independently of the environment it is a part of. ('Hey look! There's an edge on the front lawn!') I will call environments like corners and edges **sub-environments**. Other examples include regions denoted by words like *top*, *side* and *back*.

How should sub-environments be represented? They are better understood as identifying regions of an environment in a frame of reference given by the environment's own internal geometry. The key question is therefore, how is the environment's geometry to be represented?

I propose that an environment's geometry is defined in terms of its motor affordances. For a object which is small enough to manipulate, these will relate to surfaces and opposition spaces which afford particular grasp or contact relations. Thus the edge of a desk might be a region implicated in several operations involved in moving it, or lifting it. For an environment large enough for us to navigate through, the geometry largely concerns boundaries to navigation, or surfaces which provide stable support; for instance a corner is a region of an environment where two boundaries meet at a relatively acute angle, and where possible paths for the agent are particularly constrained, while the 'middle' of an environment affords many more navigation options. I therefore suggest that a sub-environment, labelled by words such as ‘top’ or ‘corner’ or ‘middle’, identifies a region of an environment by its affordances, either for navigation or for prehension or manipulation.

Note that some sub-environments can also be classified as objects. For instance, the lid of a box or the door of a room can be defined in relation to the affordance-based geometry of the environments which they belong to, but they are also manipulable and classifiable objects in their own right. (It is interesting to compare a *door* to a *doorway*: a door has a clear affordance-based characterisation in a room environment if it happens to be ‘in’ a doorway. If not, it’s just an object.)

13.4.4.3 Configurations of objects as environments

It is interesting to reconsider the hierarchical object classification system described in Section 10.5 in the light of the current discussion of environments. Some of the categories used in that section were *shapes*, such as lines and circles. If a salient region was categorised at a low spatial frequency as a shape, and then at a high spatial frequency as an object, we produced descriptions such as ‘a line of dogs’, or ‘a circle of wolves’. Note that lines and circles, just like edges and corners, cannot be established as objects in their own right. (‘Look mum, there’s a line in the garden!’) However, it is certainly possible to think of them as environments. Recall that when we establish an object ‘as an environment’, we create a saliency map of regions ‘within the object’ at higher spatial scale, to allow us to focus on a proper subpart of the object. In group classification, we move to the higher spatial scale, but without establishing a saliency map: the region classified remains the whole region associated with the object. We can think of the objects established during group classification as completely ‘filling’ the environment, or as being what the environment completely *consists of*. There are some ‘object categories’ which it is particularly helpful to think of as ‘environments’ which are ‘completely filled’ by the objects they ‘contain’, and which comprise *nothing but* their contents. ‘Shape’ categories such as ‘line’, ‘circle’ and so on are one such case. Note that we can talk about wolves as being *in a circle*, or soldiers as being *in a line*: the appropriateness of the spatial preposition *in* is significant.

It may also be possible to think of *numbers* as relating to configurational environments. Very briefly: assume we establish a group of objects as a single region in the saliency map, and then set up a sub-saliency map within this region individuating the objects in the group. Counting the objects in the group involves an exhaustive serial search, synchronised with a serial verbal routine. The point is that the counting process operates on the group object *defined as an environment*. This idea may be consistent with our usage of the preposition *of* in expressions like *three of the dogs*, or of the preposition *in* in expressions like *dogs arrived in large numbers*, or the Italian *eriamo in tre* (there were three of us). Note that a number like ‘5’ is not itself a configurational environment: each configurational environment is an ‘instance’ of a number. Numbers themselves should be thought of as *properties* of configurational environments. Some of these properties can be determined directly, by numerosity judgements (or by parallel individuation in the case of sets of four or less). More mathematical number properties can only be determined by the taught strategy of counting, in which an exhaustive visual search is accompanied by an attendant verbal routine. Note that counting has verbal side-effects but no sensorimotor ones: counting a set containing twenty-one objects does not modify the sensorimotor representation of this set, although it generates a special verbal label for it. (In sensorimotor terms, a set containing twenty-one objects has a numerosity representation which provides some indication of its size, but this is much less specific.) One interesting thing about 'mathematical' number words like *five*, *six*, *seven* etc is that they do not really denote precise sensorimotor representations: they denote the results of sensorimotor processes which have their own verbal component. It would be interesting to pursue a Kantian line of enquiry, and see whether concepts of number and geometry, understood linguistically, can be grounded in the present model of visual and spatial cognition.

[Syntactically, number words and numerosity words are determiners. What's the significance of that?]

13.5 The 'current spatial environment' representation, and how it is updated

Section 13.4 presented a model of how an agent represents the spatial relationships between the objects and environments he is familiar with. The basic idea is that any given environment is associated with a map of locations, any one of which can in turn be associated with another environment (either an adjacent or a nested one) or another object (which can of course be treated as an environment in its own right).

Having set out this model, it is useful to revisit the notion of the agent's *current* spatial environment—i.e. the LTM environment which represents 'the place where the agent currently is'. The notion of the agent's current spatial environment was introduced in Section 3.6.2.2. We can now say something about how the current spatial environment *changes* as the agent executes locomotion actions. In Section 13.5.1 I will consider how the current environment representation is updated as the agent physically moves from environment to environment. In Section 13.5.2 I will consider how relationships between environments are learned. In Section 13.5.3 I will discuss updates to the current environment representation which result from purely attentional actions.

13.5.1 Updating the current spatial environment during locomotion actions

I assume there is a medium in which all the LTM environment units compete, called the **current spatial environment** layer. Only one of these units can be active at any given time. Suppose that the agent is currently in a room, represented by LTM environment E_{office}. Under these circumstances, the hippocampal cognitive map will represent a particular collection of boundaries, with respect to which a set of places will be defined. There will be a particular set of associations between places and other LTM individuals (the objects in the room whose location is known). There will also be a particular set of associations in force between *allocentric states*—i.e. pairs of places and orientations—and other LTM environments (the environments known to adjoin the room, or to be nested within it). Say the room has a doorway which the agent knows accesses a garden: this knowledge will be represented as a link between a particular allocentric state (the state of being at the location of the doorway and facing 'out', which we will denote $S_{doorway}$) and an LTM environment representing the garden (which we will denote E_{garden}). If the agent happens to achieve allocentric state $S_{doorway}$, a top-down representation of the associated LTM environment E_{garden} will be activated. At the same time, the agent's environment perception system in the PPA will generate a holistic representation of the new garden environment which he now has access to. This representation will be independently linked to a particular LTM environment (see Section 13.2.1.1). If the agent's spatial memory is accurate, his perceptual and memory representations will converge on the LTM environment E_{garden}. In this situation, an *update* will occur within the current spatial environment layer: the currently-active E_{office} individual will be suppressed (perhaps via self-inhibition), and the newly activated E_{garden} individual will take over as the dominating representation. This transition will in turn trigger a dramatic reinterpretation of the information carried by hippocampal place cells. Firstly, the locations which they represent will be defined in relation to the shape of the garden, which may be different than the shape of the office. Secondly, a new set of associations between 'locations' and other LTM individuals will come into effect, to represent the agent's spatial memory for the location of objects and environments in this new location. These changes are what underlies the 'remapping' of place cells which is observed when an animal moves from one environment to another.

13.5.2 Learning about relationships between spatial environments

I assume that every time the current spatial environment is updated, some learning takes place, so that the agent expects a similar transition if similar circumstances present themselves. In the above situation, the agent already knows the relationship between the office and garden environments. But assume that there is another doorway out of the office which the agent has never stood in (or looked through), which leads to a room the agent has never seen. Say the agent now stands in this doorway, and perceives the new room environment for the first time. His environment perception modality will generate a representation of a room, which is recognised as new (see Section 13.10 for an account of how this is done). Accordingly a new LTM environment will be created, and the agent's current allocentric state will be associated with this new environment. (Of course, this association will be conditional on the activation of the current spatial context representation E_{office}, like all associations in spatial memory.)

In summary, every time the agent transitions from one spatial context to another, some learning takes place, which strengthens an association between the agent's allocentric state in the first spatial context and a top-down expectation of the second spatial context.

13.5.3 Attentional establishment of distant environments

Recall that an agent is also able to enter a new environment 'attentionally', by directing the environment perception modality to a subpart of his visual field. An agent can establish a distant environment by this means—for instance, a distant garden—and can also establish a distant object as an environment; both types of establishment are described in Section 13.3. What happens to the 'current environment' representation in these cases?

My proposal here draws on one of the key concepts introduced in Chapter 2: the distinction between action execution mode and action perception mode. Recall that an observer's sensorimotor experience takes the form of a sequence of operations, in which early operations set up the attentional context for later operations. The very first operation is to make a decision about whether to do something or to observe something. Making this decision establishes the way in which the observer's action and intention representation systems are connected to his sensorimotor apparatus. Understanding representations within these systems is impossible without knowing which decision was taken, i.e. what mode of connection with the sensorimotor apparatus is in force.

I now propose that the observer's spatial representations are similarly dependent on a prior decision to act or to observe. I will outline this proposal in the rest of this section.

13.5.3.1 Current observer location and current subject location

To begin with, I propose that the observer maintains *two* representations of 'current location'. One of these is constrained to represent his own location; the other can be configured to represent his own location, or the location of an object, agent or environment separate from himself. Each location representation has two components: one is a representation of 'the current environment'—i.e. a single active LTM environmnent representation—the other is a representation of 'the current *place* within this environment'—for instance a single active hippocampal place cell, or a single active point in a motor map (if the environment is a manipulable object). I will term the representations of the observer's own location the **current observer environment** and the **current observer place**. I will term the more flexible representations of location the **current subject location** and the **current subject place**. The 'subject' can be the observer himself or a separate observed entity.

Representations in the current subject environment and current subject place should be thought of as analogous to representations in the motor system. What they refer to depends entirely on which mode the observer is in. If the observer is in **self observation mode**, they are simply references to the current observer environment and current observer place. If the observer is in **external observation mode**, they refer to the environment and place of *whatever the observer is currently attending to*—in other words, to the place computed by the **external object location function** introduced in Section 13.4.2.1. These terms are intended to encompass the terms 'action execution mode' and 'action observation mode', but to extend to cases where states are perceived as well as where actions are perceived. The event which triggers external observation mode is, as before, the observer's attention being drawn to an external entity.

It is useful to revisit the 'orienting function' discussed in Section 13.4.2.2 in the light of the newly introduced definitions of current observer place and current subject place. Recall that the orienting function takes an allocentric representation of the observer's own location and orientation, plus an allocentric representation of some *other* location in the environment, and generates an orienting action which establishes the observer's attention on the corresponding location in the world. In Section 13.4.2.2 it was not clear how the representations of the observer's own location and of the attended location were kept separate. The distinction between current observer place and current subject place remedies this problem.

13.5.3.2 Candidate subject places and current subject place

Recall from Chapter 2 (see e.g. Section 2.8.3) that in the 'initial context' prior to the execution or perception of a reach action, the observer is in a state in which all objects in his environment compete for attention. The observer himself is one of the objects which competes for attention, on an equal footing with external objects in his environment. In Chapter 2 we did not discuss the attentional medium in which these objects compete in any detail—however, we did specify that it had to represent the observer as an object just like any other. The 'current subject place' representation just introduced in Section 13.5.3.1 has many of the properties needed to function as the attentional medium in which the observer competes with external objects: it can represent the location of either the observer or of an external object. However, it is constrained to represent the location of just *one* object: the one *currently attended*. It is useful to envisage an earlier representation, in which a set of environment-centred representations of object locations are activated in parallel, and which passes activity forward to a competitive layer in which a single location is selected. The latter layer represents the current subject place. The former layer we can think of as holding a set of **candidate subject places**, from which the current subject place is selected. The relationship between the two layers is illustrated in Figure 13.4. There is a one-to-one relationship between candidate and current subject places. Each

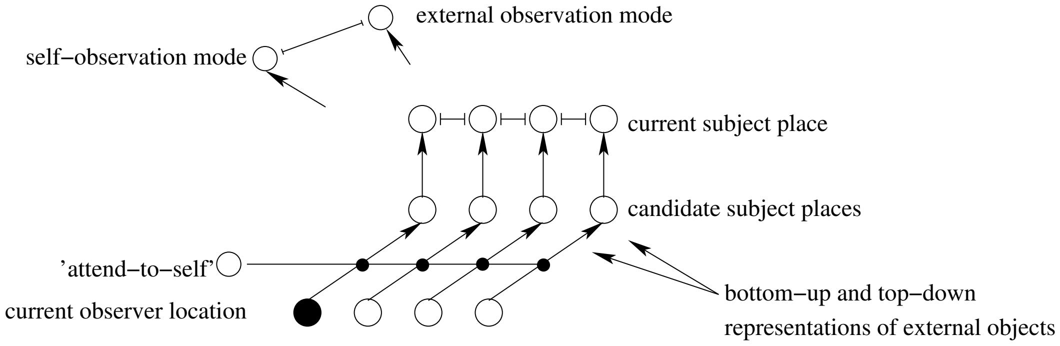


Figure 13.4: Candidate and current subject places

candidate subject place receives input from various external sources. One source relates to the observer's own allocentric location. There is a one-to-one mapping from current observer places to candidate subject places. At any time, there is a single current observer place unit active. If the agent decides to attend to himself, activity from the current observer place is gated through to the corresponding candidate subject place. At the same time, the candidate subject representations also receive input from representations of external objects in the environment. These might be bottom-up, channelled through the saliency map (via a transformation related to the orienting function), or top-down (i.e. representations of LTM individuals to be sought for, channelled through object location memory). The combined activations from all these sources are sent to the current subject place layer, in which a single place is selected. I have also sketched the circuitry which establishes self observation mode or external observation mode once a winning place has been chosen.

Note that the architecture shown in Figure 13.4 can easily be extended to support a systematic search of the objects in the current environment (including the observer), simply by adding an inhibitory link from each unit in the current subject place layer to its corresponding unit in the candidate subject places layer. This could allow the observer to search through his whole environment as well as just through just salient points in his current visual field. A sequence of current subject places will be established in turn. Each one will provide input to the orienting function, which will direct the observer's attention to an appropriate point in space. In other words, the map of current subject places can be thought of as a version of the saliency map which is stable over changes in the observer's orientation in, and even position in, the environment.

13.5.3.3 Object location memory revisited

The discussion in the previous section allows us to be still more precise about the architecture of object location memory. In Section 13.4.2.3, object location memory was conceived as a set of associations from LTM individuals to 'hippocampal place cells', gated by a representation of **context**. We can now specify that object location memory is a set of associations between LTM individuals and *candidate subject places*. The fact that they are *subject* places allows object location memory to store the location of the observer as well as the location of external objects. The fact that they are *candidate* places means that the set of LTM individual activations can deliver a set of top-down biases towards particular places in the current environment *in parallel*, which can combine with bottom-up perceptual biases to select a single place. Finally, the general notion of 'current context' can be replaced by the notion of the current subject environment. The extended network for object location memory is shown in Figure 13.5. This network shows how object location memory contributes activations to the set of candidate places shown in Figure 13.4.

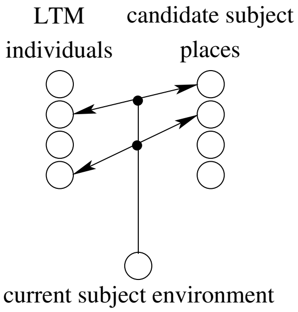


Figure 13.5: An extended model of object location memory

13.5.3.4 The possible operations in external observation mode

When the observer has decided to attend to an external entity, there are a number of things which can happen next. These will be reviewed in this section.

To begin with, the 'current subject environment' representation is set to the current observer environment representation. The entity attended to is assumed to have a place in the observer's current environment. In addition, a 'current subject place' representation is established which encodes the allocentric location of the object or location being attended to. Hippocampal 'view cells' can be thought of as the current subject place representations which are formed when the agent happens to be in external observation mode.

What happens next depends on whether the entity attended to is best classified as an environment or as an object (c.f. Section 13.3.1.1). If it is classified as an environment, an LTM environment will be associated with it. This environment is now established as the current subject environment. Since it is an environment, a saliency map of points within this environment is also automatically created. If one of these points is chosen to be attended to, its location will be given as a place within the current subject environment, rather than within the current observer environment.

If the entity initially attended by the observer happens to be an object, rather than an environment, several things may happen. Firstly, the object may be an animate agent. In this case, one possibility is that the agent starts to execute a reach action. The observer's 'current subject environment' will then change from his own environment to a representation of the observed agent's peripersonal space. (Note that this is another way of representing the second operation in the perception of a reach action, as described in Chapter 2. A full reinterpretation of the spatial representations involved in executing or observing a reach action will be given in Section 13.7.) Another possibility is that the observed agent begins to execute a *locomotion* action. In this case, the observer's 'current subject environment' remains the same as the observer's own environment, and the observer's 'current subject place' moves through a sequence of states representing the observed agent's trajectory. (A fuller account of the perception of locomotion actions is given in Section 13.11.) Another possibility is that the observer is not interested in the observed agent's actions, but wants to find out about objects which the agent 'possesses'—i.e. which are part of, or belong to, the agent. In this case, a new LTM environment becomes active, which represents the agent as an environment, and this new LTM environment becomes the new 'current subject environment'. At the same time, a saliency map is established on the observed agent. Points in this saliency map will be related to places 'in the agent', which in the first instance will be motor systems. These points might correspond to 'parts' of the agent, such as arms and legs, or to things held or possessed by the agent. (I will talk more about how objects and body parts are indexed to an agent in Section 13.7.)

Secondly, the object attended to by the observer might be an inanimate one. In this case, the range of possibilities is somewhat similar. The object may undergo an internal configuration change—for instance, a piece of paper may curl in the fire, or a door may close. The object may move through the environment—for instance, a ball may roll through the room, or a leaf may fall. Finally, the observer can decide to attend to the object as an environment. In this case, again, a new LTM environment will become active as the current subject environment, which represents the object as an environment, and a saliency map will be established on the object, representing interesting locations within it. If one of these locations is attended to, the object found there will be represented as a point in the newly established current subject environment (i.e. as an object which is part of the originally attended inanimate object), rather than as a point in the observer's own environment. For instance, the location of a cup found by establishing a desk as an environment will be given in relation to the desk, rather than in relation to the enviromnent which the desk is in.

13.5.3.5 The relationship between LTM individuals and LTM environments

In the previous section I emphasised the idea that an LTM individual can become an environment in its own right, because it has important consequences for the architecture of the spatial LTM system. I assume that the representations of individuals and environments must be physically separate, as a way of enforcing the ontological distinctions between them. In this case, the possibility of reinterpreting an object as an environment requires there to be a *mapping* between LTM individuals and LTM environments. This mapping, where defined, should be symmetrical, and one-to-one: a single LTM individual should be bidirectionally associated with a single LTM environment. Whenever an object is established as an environment, a new element of this mapping is learned.

13.6 Interim Summary

Sections 13.2 and 13.3 outlined a model of how objects and their locations in environments are perceived. This model introduces some perceptual modalities which are specialised for perceiving the spatial structure of environments. It also suggests that the duality between objects and environments relates to a hierarchical system of saliency maps. The duality between objects and environments in the perceptual model is echoed in the model of spatial LTM, by a device which maps LTM individuals onto LTM environments.

Sections 13.4 and 13.5 outlined a rough model of spatial LTM, and of how it is used to represent the environment which an observer is currently in, or currently attending to. The basic proposal is that individual objects and environments in the world are represented as individual neural assemblies: objects are represented by a collection of LTM individuals, and environments are represented by a collection of LTM environments. I have defined a notion of the *current subject environment*; this representation can be used to encode the observer's own environment, or an environment which is distant from the observer. I have also defined a notion of the *current subject place*; for any current subject environment there is a set of these places, which are defined in relation to the environment's boundaries (for the kinds of environment which are navigable by an autonomous agent) or in relation to the haptic affordances of the environment (for the kinds of environment which are explored by motor subsystems such as hands, fingers, and other effectors). I have also defined a way for the spatial relationships between environments and between objects and environments to be encoded in LTM; the basic proposal is that a given current subject environment enables a set of associations between place representations and other objects or other environments. I have also described how these relationships are learned and used by an agent to generate top-down expectations.

13.7 Spatial representations in the reach-to-grasp action

Before moving on, it is useful to reconsider our cup-grabbing action within the model of spatial representations which has just been given. There are several interesting connections to be made. I first consider how an agent should be represented in spatial LTM. An agent is an object; therefore it should make sense to be able to establish an agent 'as an environment'. In Section 13.7.1 I show that this fits well with the sensorimotor model of agents developed in Chapter 2. I introduce a new kind of environment called a **motor environment**, and explain how this can be incorporated into the model of objects and environments already developed. In Section 13.7.2 I consider how a manipulable object should be considered as an environment. This creates links back to two concepts introduced in Chapter 2: one is the notion of 'opposition spaces', and the other is the notion of 'haptic' perceptual modality. Both these concepts can be integrated within the general model of spatial cognition just developed. In Section 13.7.3, I discuss the special mode of motor control which allows agents to bring about various types of movement in a grasped object. Finally, in Section 13.7.4, I look at how to represent the establishment of a stable grasp in spatial LTM. The basic proposal is that the object should become indexed to a position in the agent's motor environment, and by this means ultimately linked to the agent.

13.7.1 Agents as environments

Recall from Section 13.2.2 that there are several spatial representations which are specific to agents. Some of these are holistic representations of agents—these might be representations of individual agents computed in the object classification system, or representations of global poses or motion gestalts computed in the somatic or biological motion systems. If an agent is established 'as an environment', then the dominant representation of places 'within' this environment will not be given by hippocampal place cells; rather it will be given using motor representations, as discussed in Section 13.2.2.2. I will assume that there are two different *types* of environment, with separate systems for representing maps of places. A **gravity environment** is predominantly horizontal, and affords navigation by an autonomous agent who can maintain contact with it due to the force exerted by gravity. Places within such an environment are given in the hippocampus. A **motor environment** is an agent, and places within the agent are defined using motor representations.

What are the places within an agent? The motor system is strongly hierarchical, as shown in Chapter 2. Accordingly, there are many levels at which places can be defined. As already suggested in Section 13.2.2.2, I propose that at the highest level, the map of places within an agent is defined as the set of motor systems which the agent can deploy to reach points in his perispace—for instance, an arm/hand, or a leg/foot, or a pair of arms working in coordination. I assume that these systems compete amongst one another, and that when an agent decides to act, one system is selected.

When one motor system is selected, I suggest that it becomes established *as an environment in its own right*; in other words as a context within which a finer-grained collection of locations can be represented. If the selected motor system is the hand/arm, then a new motor map can be referenced, representing the objects in the vicinity of the selected arm as movement vectors. When one of these objects is selected, we again establish a new environment: places within this new environment are represented using a motor map centred on the hand, which is a motor subsystem of the hand/arm. At this point—at least, if the goal is a stable grasp—the hierarchy bottoms out.

The notion of **sub-environments**, introduced in Section 13.4.4.2, is useful for describing hierarchical relationships within motor systems. I propose that the hand is a sub-environment of the arm; in other words, that the hand is a *place* within the arm, which is defined in relation to the geometry of the arm. Just as some sub-environments can also be characterised as objects in their own right, we can treat motor systems like the hand and arm as objects; however in their role as motor systems, they are best understood as sub-environments.

On this model, if a motor system is seen as an environment, then 'attentionally entering' a motor system basically means selectively activating one of its component subsystems. The active subsystem can be thought of as the experiencer's 'place' within an established motor environment.

13.7.2 Target objects as environments

In Section 13.3.5 I introduced the idea that the spatial structure of a manipulable object established 'as an environment' is given by the types of action it affords to motor systems with effectors of the appropriate scale. In this section I will elaborate on this idea.

The basic suggestion is that 'places' within an object established as an environment are defined as actions performable by an appropriate motor subsystem. Some of these actions may be stable grasps. But other actions may be more exploratory—for instance, the running of a finger around a contour, or the running of a palm or set of fingers over a planar surface. As already discussed in Section 2.5.3, areas like AIP compute an object's affordances from visual information. Once these areas have learned their task, an object can be established 'as an environment' *before* it is reached by the hand, and even before it is in the agent's perispace.

When an agent is reaching for an object, there must be a mechanism which *selects* a particular contact or grasp action from amongst the different actions afforded. I now assume that there is a visual attentional component to this mechanism, so that when a particular grasp or contact action is selected, certain *places* within the region subtended by the cup are also attended to: these will be the landing points or surfaces for the action.

Now recall from Section 13.3.1 that when an object is established as an environment, a sub-saliency map is also created within the region it subtends, so that objects 'within' this environment can be attended to. Active regions in this saliency map are regions occupied by objects. There must be a mechanism for associating regions in this map with places defined within the agent's motor vocabulary. I assume that the visual attentional mechanisms which are correlates of action selection are also expressed within the sub-saliency map—in other words, that the saliency map provides a mechanism for expressing the location of objects in 'places' on their 'host' object. Thus, for instance, if I notice that a cup has a fly on its rim, the fly is an object, and the rim is the place at which the object is located. The rim is ultimately defined in terms of its motor affordances, but because it is also expressed as a region within the sub-saliency map, we can also represent objects which happen to be located 'on' the rim.

Note that we expect many regularities in the relationship between objects and the host objects they are 'in', or 'on'. These mainly relate to relationships of stable support, or contact. Places on an object are ultimately defined in terms of these relationships, and we expect objects 'in' these places to have appropriate configurations. For instance, we expect the fly to be 'standing' on the rim of the cup, not lying on its back, or attached by one wing. These correspondences presumably have to be learned through long experience with combinations of objects during infancy.

13.7.3 Manipulated objects as observed agents

A stable grasp is not always the ultimate goal of a prehension action. When a target object is held in a stable grasp, the agent can control it: movements of the arm will change its location, movements of the wrist will change its orientation, and if it is soft or flexible enough, movements of the fingers could cause a change in its internal configuration. (In fact, changes to the object can also be brought about by many types of contact which do not involve a stable grasp, as for instance with pushes or punches.) Recall that in order to achieve contact with an object, the agent has to learn a *sequence* of three sensorimotor operations: attention-to-self, attention-to-target, motor-action-execution. Note that these actions can be thought of as actions executed within increasingly nested sub-environments: the first action takes place within an allocentrically defined environment, in which the agent himself competes with external entities for attention; the second takes place within the motor environment of the agent; the third takes place within the motor environment of the arm. We can now postulate that in actions defined by the effects they bring about on a target object, a fourth operation must occur, which transitions to a *fourth* attentional environment. Interestingly, this new environment is not a sub-environment of the hand. I suggest instead that the agent must attend to the manipulated object *as an agent in its own right*, using the perceptual systems he would employ to attend to an external agent performing some arbitrary action. I propose that the agent enters a new mode of motor control, in which feedback is provided by the action recognition system, rather than by specialised mechanisms for monitoring reach actions. There are several interesting aspects to this idea.

Firstly, note that during this new mode of attention to the target object as an agent, it is not necessary that the agent maintains a stable grasp of, or even contact with, the object. A sequence of actions which results in the object being propelled to a distant point could quite feasibly be learned: the final attentional context in this case will involve monitoring the object flying through the air, using the same perceptual mechanisms by which an external agent's locomotion actions are tracked. We can thus model actions which bring about a very wide range of effects on the target object.

Secondly, note that when the agent has established the manipulated object as a locomoting agent in its own right, he is also likely to switch to a different *environment* representation, commensurate with the size of the object. Establishing a manipulated object as an agent provides a way the agent can 'enter' environments which are an order of magnitude smaller than the environments he walks around in.

Finally, note that the agent can also establish *his own effectors* as agents, simply by executing the special fourth action without having grasped anything. This gives the agent a special way of attending to 'his own actions' with particular deliberation, which might be useful when new actions are being learned.[6]

13.7.4 Representation of a stable grasp within spatial LTM

The agent and the cup are both objects in the environment, and they must both have locations. If the cup is in the agent's perispace, and thus close enough to reach, it is likely that the agent and the cup are represented in the current environment-centred cognitive map as being 'at the same place'—at least at some granularity of resolution in this cognitive map. How is the location of the object represented once the agent has established a stable grasp?

13.7.4.1 Indexing of the object to the agent

While the agent has a stable grasp on the object, the location of the object can be read off the location of the agent. This axiom is what permits the learning of the mapping from

---

[6] Maybe babies learn their initial hand actions via this mechanism. They certainly spend a lot of time looking at their hands, unlike mature adults, who have better things to look at.

visual to motor representations of space and orientation in Section 2.5.4. But note that the same axiom must prompt a change in the way the allocentric location of the cup is specified in spatial LTM. Note that as soon as a stable grasp is achieved, I can say that the cup is *in the agent's hand*. While the agent's hand was approaching the cup, it made sense to think of the cup as the environment, and the hand as a locomotor moving into this environment. When describing the *shape* of the cup, it still makes sense to think this way. But when describing the *location* of the cup, we must now flip this relationship on its head: the agent's hand is now the *environment*, and the cup is an object contained within it.

Now recall that the agent's hand is itself a sub-environment within the larger motor environment which is the whole agent. So the spatial relationship between the agent and the held object is given by the relationship between the relevant motor environments.

13.7.4.2 Physical possession

When a stable grasp has been achieved, the LTM individual denoting the cup must be *indexed to* the LTM individual denoting the agent. From this point on, in order to re-attend to the cup, we must first establish the agent as a motor environment, and then select one 'place' within this environment, i.e. one of the agent's motor subsystems. This establishes an attentional environment in which we can directly access the cup.

13.7.4.3 The representation of body parts

Note that we can establish body parts as objects as well as as motor systems. This process is at the heart of the mapping which is established between motor and visual modalities. But note also that there is an interesting difference between saying *I have a hand* and saying *I have a cup*. The latter sentence basically functions to index the cup to a particular agent (me). It is unspecified 'where' the cup is in the agent's motor environment, but this extra information can easily be provided: *I have a cup in my hand*. But *I have a hand* cannot be thus modified. *I have a hand* has a status similar to *The cup has a rim*, or *The tabletop has an edge*.

13.7.4.4 The abstract concept of possession

Note that we can index an object to an agent when it is within the agent's perispace, even if the agent is not currently grasping it. The important thing is that the agent has some measure of control over the object, because he is able to grasp it if he so wishes. Thus we might recognise that John 'has' a big stick, even if the stick is simply lying next to him. Using this coordinate-system-based conception of 'have', to have is not necessarily to hold.

It is not a long step from here to a more abstract notion of possession. The important commonality is that objects can remain indexed to an agent *in some long-term memory domain* even when the agent moves away from them altogether. The notion of 'control' in the more abstract case has more to do with claimed rights over objects than with immediate accessibility. This abstraction from spatial proximity to possession has been quite widely discussed in cognitive linguistics (see e.g. ??). My main interest is in how the more abstract notion of possession can be implemented within the LTM framework which is currently being developed.

In experience mode, establishing an object indexed to an individual involves establishing that individual first—this creates an attentional context in which the indexed object becomes accessible. I suggest that the abstract notion of possession involves operations in *memory mode*, which navigate the indexing relationships between LTM individuals in a way which parallels sequential actions of attention. I assume that the sensorimotor operation which establishes the currently attended object as an environment by adopting a finer spatial scale has an analogue in memory mode. The currently attended object is the currently active LTM individual. The memory mode analogue of establishing an object as an environment is the operation of activating all the LTM objects *directly indexed* to the currently active individual as candidates for the next LTM object to be activated. Again, the details of this operation will be given later, in Section **??**.

Note that both the sensorimotor and the memory-mode conceptions of possession involve sequences of attentional operations. An individual is established indirectly, by first establishing the individual which owns it and then performing a particular type of attentional operation. In Section **??** I will propose a mapping between sequential structures of this form and the LF structure of possessive DPs.

13.8 Simple 'hand-scale' environments

In Section 13.7.3 I introduced the idea that manipulated objects are like agents, moving around in their own environments, which are 'smaller' than the agent's own environment. A simple example of the kind of environment which a manipulated object can move around in is a tabletop: something that supports objects. Children have to learn the properties of such environments; see e.g. Baillargeon (1994; 1995; 1998). In the first instance, this learning doesn't involve manipulated objects at all - rather it just involves the infant's own hands. In this section I will outline a model of how infants learn about simple flat surfaces like tabletops, which can serve as environments within which their hands can be given trajectories. The basic aim is to explain simple actions like moving one's hand *onto* a tabletop, or *along* it.

The account is analogous to the account of learning about objects (and how to reach/grasp them) given in Lee-Hand and Knott (2013). In that account, the infant begins by finding certain tactile stimuli intrinsically pleasurable, and this leads to the development of a simple motor conception of an object, as a stable grasp state. Then the infant learns the visual representations which map onto these motor representations, so s/he can recognise objects (and their grasp affordances) visually. I suggest that the account of how infants learn surfaces runs parallel to this, but that it's nonetheless quite separate: it involves learning in a 'space perception' modality, rather than the object perception modality. As discussed in Section **??**, these are quite different.

For bootstrapping the concept of a surface, I assume the basic axiomatically pleasurable tactile sensation is a sense of **even touch**. The sensation involves tactile responses from several parts of a single hand - in this discussion, I'll assume these are the tips of all five fingers. The pleasurable sensation is a touch of a particular intensity. When this same intensity of touch is felt in all the fingertips, it must be the case that the infant's hand is **aligned** with the surface. There is a plane defined by the fingertips which is aligned with the plane of the surface being touched.

The next axiomatically pleasurable tactile sensation is a *slip*. The human hand has slip sensors (citation). While slips must be avoided in the establishment of a stable grasp, for the system which learns about surfaces they are positively rewarding: they encourage the exploration of surfaces.

I suggest that the infant is axiomatically encouraged to make movements which generate slip sensations, and which also preserve evenness of touch. The infant moves his hand about at random, but learns how to alter his wrist orientation (or perhaps finger positions) so as to correct uneven touches over his/her fingertips, and how to make corrections to the hand trajectory so as to maintain a suitable absolute level of touch. (There are some movements which pull back if contact becomes too forceful, and others which move forward if contact becomes too light, or is broken altogether.) After this learning, the infant can explore a surface by touch, in a mode where his fingertips are constrained to align with the surface.

At this point, visual learning can occur. Assume the child has a visual environment perception modality which maps the distance and local orientation at each point in a region of the retina onto the hand position and orientation needed in order to achieve even touch on that surface. This modality is entirely distinct from the modality which maps the visual stimulus at a given retinal location onto a grasp affordance. The environment perception modality supports *navigation* of the hand around the surface which projects onto the region of retina, rather than grasping of an object at that point, so the retinal region is likely to be larger.[7] However, the visual learning can happen in the same kind of way. At each point when the infant attains evenness of touch, during a period of time when he is sliding his fingers around a surface, a function can learn to map the distance of the visual stimulus at the point where his fingers are, and the local orientation of the surface at that point (as computed visually) onto the position and orientation of his hand. While training the function happens one point at a time, the function should be assumed to deliver these values in parallel, so that after training, the values are delivered across the whole retinal region.[8] We can now define hand trajectories in relation to the plane: wherever the hand is, we can search visually adjacent regions to find which ones we can 'slip' our hand to. Our movements to ensure even touch around the plane can then be

---

[7] At least until we start to think about 'attentionally entering' distant regions, as discussed in Sections 10.8.3.3 and 13.5.3.

[8] That's implausible! I need to go into more detail about the set of cascading functions that deliver possible hand orientations. The first function delivers a set of shoulder joint rotations, from which one can be selected; the next one takes the current shoulder joint rotation and delivers a set of candidate elbow joint rotations, from which one can be selected; the next one takes the current shoulder and elbow joint rotations and delivers a set of candidate wrist rotations.

informed by vision, i.e. by a look-ahead component, rather than just by feedback.[9]

In fact, this look-ahead can run some way in front of the actual position of the hand. Say we want to reach some particular goal point on the plane. I assume the goal position will be established using regular object-based attention: e.g. a visual location will be mapped into a goal motor state. Rather than moving directly towards this motor state, as in normal reaching, we search the local visual environment around the hand, and find the motor state which moves us closest to the goal, and proceed to that state. At this point, the hand will be moving through a succession of goal points: there is a proper notion of a 'planned trajectory' (which might even correspond to something visual, represented as a path through the plane).

As well as moving *through* a planar environment, the hand can move *into* one. This is somewhat like a reach, except that the goal is not an object that is grasped, it is a plane which the hand comes into (or into contact with). For visually-guided reaching, each point in the plane maps onto a goal hand state, so the agent can reach to any point. The one they actually reach for will be a matter for competition: presumably the easiest point to reach for will win.

Once we know about hand-scale environments like surfaces, and we know how to grasp objects, we can start to learn how to move grasped objects in relation to hand-scale environments. (Since we can grasp these objects, they move in the same scale environments as our hands.) Objects have stable configurations with their planar environments: they are supported by them. The relationship of support is analogous to the relationship of even contact: there is a planar surface of the object which is aligned with the surface which supports it. When we put a cup *on* a table, one thing we have to do is get the alignment right. We can *feel* when it's right, because it'll *feel* the same as an even touch.[10] When we explore different wrist positions when grasping a cup resting stably on a table, the effect will be the same as exporing different wrist positions when our hand itself is on the table. (Because the cup is solid.) So within the realm of touch, we can learn when a grasped object is **stably supported** by a surface. Subsequently we can learn what visual features of objects and surfaces are indicative of stable support, so we can orient an object in advance. But I expect there's normally a small haptic test when you put a cup on a table—especially if it's bone china.

---

[9] This also needs expanding. The *local* constraints on a continuous plane are a set of functions which take the full set of current joint angles, plus some movement direction (e.g. 'left', 'right', 'forward', 'back', 'up', 'down') and specify what *changes* in joint angles are necessary. There are also constraints defined in relation to surfaces on the hand: e.g. if the hand surface in contact with the surface is curved, then moving along the curve will involve rotating the wrist in the plane of the curve.

[10] Note: when we're grasping an object, we have an even touch on the object! At least within the parts of the hand which create the opposition axis. We must do, because the forces oppose one another.

13.9 Temporal contexts and the object location function

Note that object location memory as defined above tells us where objects are *at particular times*: it is a function which takes a spatial context *and a temporal context*. We have already made extensive use of the idea of 'spatiotemporal contexts' in our account of episodic memory in Section 3.6.2. However, we did not discuss at that point how to decompose spatiotemporal contexts into independent spatial and temporal components. Now that we have introduced a model of 'purely spatial contexts', we need to consider how 'purely temporal contexts' are defined. In Section 13.9.1 I will sketch a model of temporal contexts. In Section 13.9.2 I will consider some operations which abstract over temporal contexts in location memory.

13.9.1 Outline of a model of temporal contexts

How can 'times' be referred to within the memory system? Episodic memory links events to a spatiotemporal context; i.e. to a time and a place. The place can be represented as a (current) spatial context. How should a temporal context be represented?

I propose that a central concept should be an **individual situation**. Just as long-term memory stores individual objects and places, it should store individual times, or 'moments'. Each individual situation only occurs once, because time runs in one direction only. Thus when the agent is in experience mode, he is constantly creating and activating new individual situations to associate with the events and states currently being experienced.[11] Thus an individual spatiotemporal context is in fact a combination of a spatial context (which can be 'revisited' any number of times) and a unique temporal context (which can only be 'revisited' by being reactivated in 'memory retrieval mode', as discussed in Section 3.8.2.4).

A key issue concerns how frequently a new individual situation is created. A maximally precise model is to create and establish a new individual situation after every event which is experienced.[12] However, this model is inefficient in cases where a sequence of events commonly reoccurs. I will revise it in Chapter 17, after a working memory concept of **situation types** has been introduced. This concept will permit us to formulate a model of hierarchical structure for individual situations, and a general model of abstraction over times.[13]

[This should also be the place to prefigure the idea that the hippocampus creates individual situations very often, as a way of orthogonalising experienced events, but that the number of individual situations is reduced during consolidation of hippocampal memories in longer-term storage.]

---

[11] The idea of a constant stream of new individual situations is one way of understanding the 'orthogonalising' operation which occurs early in many models of hippocampal memory—see e.g. Rolls (1996).

[12] In logical models of knowledge representation, this way of representing times is used in a well-known formalism called the **situation calculus**.

[13] Once that's been introduced, we can also discuss how language allows us to refer to individual situations (at different levels of hierarchy) using *proper names*, as well as by evoking their content.

13.9.2 Abstractions over temporal context in object location memory

Many aspects of our environment remain relatively stable over time. For instance, the configuration of rooms in a house or of streets in a city tends to be invariant from one situation to the next. This can be exploited. Recall that the object location function takes a spatial context *and a temporal context*, and returns a set of associations between objects (or environments) and locations in the cognitive map. But I often want to know about the location of an object or environment *now*, based on information I gained some time before.

I assume that for objects or environments whose location is relatively unchanging, the object location function learns to ignore the temporal component of its input, because this changes without affecting the result. The spatial indexing of environments in relation to one another is likely to make extensive use of this kind of abstraction over time. For objects, the degree of temporal abstraction depends on how frequently they change location.

How do I represent the location of an object which is known to move around a lot? There are likely to be two general ways of doing this. Firstly, there may be a number of locations at which the object frequently appears. (Think of a hairbrush in a busy household.) Temporal abstraction will establish a prior distribution of possible locations activated to different degrees. Secondly, given that objects must maintain spatiotemporal continuity, there should be a bias on this prior distribution towards recently-established locations. Thus if the hairbrush was recently in the bathroom, 'it can't have gone too far' (a sentiment often expressed by people looking for things).

Recall that we are currently discussing *allocentric*, *long-term memory* representations of object location. Of course, these are complemented by egocentric representations, in working memory and sensory modalities, which also encode assumptions about the spatiotemporal continuity of objects. The bias of current object location representations towards recently established locations provides continuity between working memory and long-term memory representations of location.

13.10 Recognition and categorisation of individual objects and environments

The recognition of an individual object or place is a fundamental cognitive operation. When we recognise an individual object or place, we align the atomic components of our representation of the world with things in the world itself. In order to recognise an individual in the world as a particular LTM individual, two conditions have to be met. Firstly, there has to be a reasonable correspondence between the current *properties* of the world individual (as delivered by perceptual mechanisms) and the properties of the LTM individual (as delivered by memory mechanisms). Equally importantly, there has to be an appropriate relationship between our current *location* and the location associated with the LTM individual.

Even though our model allows objects to be established as environments and vice versa, it is useful to use a more commonsense notion of objects and environments when discussing how they are recognised. I will define 'commonsense environments' as objects whose location never changes and whose properties are perceived by the scene perception modality, i.e. in PPA. And I will define 'commonsense objects' as objects whose location *can* change, and whose properties are perceived by the object categorisation modality.

13.10.1 Recognition of individual objects

As follows:

0. The most recent memory of the object is the one which drives all of the following. (We don't use a probability distribution any more.)

1. The object as currently perceived doesn't have to have identical properties to those most recently remembered. But they have to be close enough.

2. It has to be plausible that the individual has travelled from its last-known location to the current location in the time available.

3. Recognition is based on a mixture of the above two criteria.

4. Give a definition?

5. We can get into abductive explanations, but that would take us too far afield.

13.10.2 Recognition of individual environments

It looks like the PPA is involved in both categorisation and recognition (...)

The PPA does not respond to objects, even if their semantics are strongly diagnostic of a particular type of scene.

There are two additional points I want to make.

What mechanisms are involved in landmark classification? I assume that a classification of large surfaces and their orientations is one important part of its operation. Mechanisms for computing orientation from texture or binocular disparity gradients in the attended-to region are thus likely to be involved (see e.g. ??). Straight texture classification will also be useful. For instance, grass, earth, water, or rocky surfaces can identified by a characteristic texture.

Secondly, there must be strong correspondences between the landmark categorisation system and the *object* categorisation system. Recall that objects are categorised in two ways: a representation of object type is computed in inferotemporal cortex (see Section**??**), and a representation of object *shape* is computed in AIP (see Section **??**). I assume that both these modalities are connected to the landmark categorisation system. If a 'place' is far away, it is likely to be established using focal attention, and IT-based object categorisation. For instance, a house is classified very differently if it is a distant object or if one is in it. From a distance, [14]

---

[14] The flipside of this is that it should be possible to classify very large objects (or objects which one is very close to) using the landmark categorisation system mapped into the object representation system.

One of the reasons why these correspondences are important is to provide the abstractions which are needed to recognise a locomotion action regardless of whether the agent is onesself or a third party. We must be able to recognise the action of 'entering a house' whether we are experiencing it as the agent or watching someone else do it. In the former case, 'the house' will be established in the scene representation modality, as an environment which encompasses the whole visual field (and beyond). In the latter case, 'the house' will be represented as an object, and the observer must track the agent's path 'into' the object in an object-centred coordinate system.[15]

13.11 Execution and perception of locomotion actions

In this section I present a model of how an agent can use his knowledge of locations in the world to navigate through it, or to recognise movements through the world by other agents or objects. Presenting this model also requires a number of extensions to the model of spatial LTM presented in Section 13.4, since concepts of space and concepts of movement through space are very closely related.

Section 13.11.1 introduces the concept of a locomotion action. Section 13.11.2 discusses the different types of goal which can be pursued by locomotion actions. Sections 13.11.3 and 13.11.4 outline some of the important spatial representations involved in the planning and control of locomotion actions. Section 13.11.5 outlines a model of how an agent generates and controls a locomotion action through his current environment. Section 13.11.6 examines the commonalities between the execution of locomotion actions and the recognition of locomotion actions being performed by other agents or objects. Section 13.11.7 presents a schematic model of the perception of locomotion actions.

The model of locomotion actions introduces some new elements which help to clarify the general model of spatial LTM. Section 13.11.8 extends the object location function to integrate it with the model of locomotion actions. Finally, Sections 13.11.9 and 13.11.10 outline a model of navigation *across environments*, which draws on the newly extended function.

13.11.1 Locomotion actions

I will begin by introducing the concept of a locomotion action—in other words, an action which moves the agent through his environment.

There are some interesting similarities between locomotion actions and reach actions. Firstly, we can in each case talk about a trajectory associated with the action. For a reach action, the object moving along the trajectory is the agent's hand; for a locomotion action, the object describing the trajectory is the whole agent. Secondly, we can in each case talk about an entity in relation to which the trajectory is defined. For a reach action, the entity is the target object. For a locomotion action, the entity must be defined more generally. For one thing, locomotion need not be *to*, or even *towards*, the entity. It can

---

[15] Perhaps a bit more here.

be away from it, or around it, or past it. Moreover, the entity in relation to which the trajectory is defined is not always an object—it can be an environment in its own right, whose own spatial properties are what guides the agent's locomotion. I will use the term **landmark** to describe the object/environment used to define the agent's trajectory. (The term is intended to encompass both objects and environments, picking up on the duality introduced in Section 13.3.) Some examples of landmarks are useful, to illustrate the range of possibilities. Consider an agent in a room. The agent can move *through the room*, or *along the floor*: in this case, the landmark is the environment in which the agent is embedded. The agent can move *to the desk*: in this case, the landmark is an object *in* the room, which defines the agent's trajectory by virtue of its location. The agent can also move *under the desk*: in this case, the landmark is initially an object, but once the agent is close enough to it, it is re-established as an environment in its own right, and the agent's final movements are *reconfigurations* necessary to enter this environment (see Section 13.4.3.2). The agent's trajectory during these reconfigurations is determined by these reconfiguration actions, just as the trajectory of the hand onto a grasped object is determined by the requirements of a particular grasp. Landmarks can also be 'special environments' such as paths or sub-environments, as discussed in Section 13.4.4. The agent can move *along a path* (which means to follow a particular trajectory in his current environment), or *into a corner* (which means to move to a particular location in his current environment). Even 'configurations' can be landmarks: a single agent can move *into a sitting position*, or a group of agents can move *into a shape* (such as a line or a circle).[16]

As well as a defining landmark, locomotion actions have a **means** and a **trajectory type**. The means refers to the method of locomotion: for instance walking or running. The trajectory defines the path taken by the agent in relation to the landmark. We have just seen that landmarks can be environments or objects or both; there are a range of trajectory types which reflect these different cases. A trajectory like *through* requires that the landmark is the agent's current environment, and simply identifies one of the paths currently afforded by this environment as the one travelled by the agent. A trajectory like *to* requires that the landmark is an object in the agent's current environment, and identifies a path afforded by the environment which brings the agent to the same location as this object. A trajectory like *onto* has two stages: in the first stage the landmark is treated as an object, and the trajectory identifies a path which brings the agent to the object; in the second stage the landmark is established as an environment, and the trajectory identifies a **goal configuration** of the agent in relation to this new environment. The attainment of this goal configuration is likely to influence the agent's final trajectory as he approaches the landmark.

There are many parallels between these latter trajectory types and reach action categories. Recall from Section 2.6.1 that a reach action category is defined as a characteristic hand trajectory and a characteristic goal configuration of the hand in relation to the target

---

[16] Maybe even *groups* can be landmarks. For instance, an object can smash or split or break *into [some number of] pieces*. Since this stretches the intuitive notion of locomotion further than it normally goes, I will not consider this type of landmark in the current discussion.

object. The latter stages of the hand trajectory are influenced by the goal hand configuration. The trajectory of a locomotion action is similarly a mixture of a path towards a location, and an establishment of a goal configuration which may influence the final stages of the trajectory. Also, recall from Section 2.5 that a reach action involves a shift in coordinate systems. In its early stages it is primarily about transporting the hand close to the target. In its latter stages, it is about achieving a goal configuration of the hand in relation to the target, namely a stable grasp. For a locomotion action which involves a goal *configuration*, we can similarly think of control switching to a coordinate system centred on a newly established environment.

In summary: I will assume that a locomotion action is defined by a **landmark**, a **locomotion type**, and a **trajectory type**. The *classification* of a locomotion action involves establishing the locomotion type and the trajectory type. (The locomotion type can be established with different levels of precision, with *go* being the most general category.) To look ahead to connections with language: motor action types will correspond to verbs of locomotion, and trajectory types will correspond to a subset of the **spatial prepositions**—specifically, those which define a trajectory (and optionally a goal configuration).

Note that a single locomotion action might be classified in several different ways, using different pairs of landmarks and trajectories. For instance, if an agent in a room moves to the door, we could describe the action as movement 'to the door', but also as movement 'through the room', 'over the floor', 'past the desk' and so on.

13.11.2 The goals of a locomotion action

Every voluntary locomotion action is presumably executed in pursuit of some goal. When we describe such an action, it is important to do so by referring to the agent's goals—recall from Section 2.7.6 that the deepest descriptions of actions are at the level of underlying intentions. For instance, if a man in a park is chasing a dog, his goal is to get *to the dog*—his action may also take him 'away from' a particular tree, but this is only coincidentally true. In this section I will consider several issues which arise in describing the goal of a locomotion action.

13.11.2.1 Two types of locomotion goal

Some locomotion actions have as their goal the attainment of a particular location in the agent's current environment. I will call such goals **location goals**. I suggest that locomotion actions have location goals when the landmark for the locomotion action is an object in the agent's current environment—i.e. when the agent and the landmark both have positions in the current environment. The most obvious location goal is to *reach* the landmark. There are other cases where the goal is to achieve a specific location *relative to* the landmark—for instance, a location 'beyond' or 'past' it.

Other locomotion actions have as their goal the *exploration* of the agent's current environment. For instance, an agent might move around a room or a house in order to find out what it contains. I will call such goals **traversal goals**. I suggest that a locomotion action has a traversal goal if the landmark is the agent's embedding environment. The trajectories described during such actions are simply those trajectories which are afforded by the environment. We might imagine an inhibition-of-return mechanism operating over these trajectories, to encourage an exhaustive search of the environment. An interesting special case is when the agent is in a 'path environment' (see Section 13.4.4.1). In this case, exploration just consists in following the path. The path will of course lead to a location—but if the agent's primary goal is to traverse the path, attaining this location is not his primary objective. Consider an athlete, whose goal is to run round a circular track. Here the end location is the same as the start location, so clearly the locomotion action cannot be motivated by the attainment of a location.

Finally, there are locomotion actions whose goal is defined negatively, as being 'away from' a certain landmark. The landmark could be the agent's current environment, or something else embedded in the environment. There is no specific goal location for actions of this kind; the requirement is simply the attainment of a certain distance from the landmark.

13.11.2.2 Hierarchical structure in locomotion goals

An agent often has a hierarchical structure of locomotion goals. Consider an agent who wishes to get *to his office door*. It may not be that he can achieve this directly: there may be an obstacle in his way. The agent must first navigate *round the obstacle* in order to achieve his main goal.

It is useful to define an **obstacle** as a landmark which the agent must navigate round, or past, in order to achieve some higher level navigational goal. We can define a special class of locomotion action, where the landmark is an obstacle. For such an action, the agent has a location goal—the location to be reached is one at which the obstacle no longer blocks the agent's path to the target. Of course, when the agent navigates past one obstacle, he may encounter another. An agent's navigation to a goal location can often be thought of as a sequence of obstacle avoiding environments.

It is also useful to think of navigation *through the current environment* as a subgoal. When an agent wants to reach a goal location, he must navigate through a certain portion of the environment in order to achieve this. Note that the environment only affords trajectories which avoid obstacles. So we can think of subgoals negatively, as navigating round obstacles, or positively, as following paths in the environment.

When we are describing a locomotion action, we can use different levels of granularity. We can refer to the agent's ultimate navigational goal, or to one of the agent's subgoals. We can also refer to several goals at once: for instance a locomotion action could be described as going 'down the corridoor to the fire exit', or 'round the defender to the try line'. The important thing is that the description picks up on *some* aspect of the underlying intentional structure.

13.11.3 Allocentric representations involved in the planning and control of locomotion actions

In the following two sections, I will outline some of the important neural representations involved in the planning and control of locomotion actions. In this section I will consider allocentric (environment-centred) representations, and in Section 13.11.4 I will consider egocentric representations.

In the current section, I will first discuss evidence that trajectories are a natural unit of encoding in the hippocampus. I will then discuss neural representations of goal locations, and outline a model of goal selection and trajectory selection, drawing on a mixture of look-ahead and online mechanisms.

13.11.3.1 Hippocampal representation of trajectories

There is mounting evidence that hippocampal ensembles in rats can encode *sequences* of locations, as well as individual locations. As already mentioned in Section 3.7.1, Frank *et al.* (2000) found that the firing of a hippocampal place cell sensitive to a given location was modulated by the path the animal had taken to arrive at that location, and also by the path the animal subsequently took. Ferbinteanu and Shapiro (2003) found similar prospective and retrospective sensitivity to the animal's trajectory in hippocampal cells. In addition, they found that prospective encoding was diminished in trials where the animal made a navigational error, suggesting that this type of encoding is involved in the animal's goal-oriented navigational behaviour. Most recently, Ji and Wilson (2008) studied rats in the process of switching from a well-learned trajectory to a new, partially overlapping trajectory. They found that learning the new trajectory caused the activity of cells associated with the overlapping region to become increasingly sensitive to past locations on the new trajectory. This suggests a mechanism for learning a new trajectory, involving the creation of new dependencies between cells encoding successive positions on the trajectory. Ji and Wilson also found that the changes associated with new trajectory learning preceded a reliable behavioural switch to the new trajectory—again, this suggests that the changes have a causal role in influencing the rat's behaviour.

In summary, there is good evidence that trajectories are a natural unit of information encoding in the hippocampus. Clearly, for some purposes, the hippocampus must represent static locations—for instance, object location memory requires this form of encoding. I will assume that the hippocampus can naturally store both static locations and trajectories within the agent's current environment.

13.11.3.2 Neural representation of goal locations

While the neural representations of an agent's current location are quite well understood, the neural representations of an agent's goal locations have been harder to discover. Prospective encoding of the agent's trajectory presumably *reflects* the agent's ultimate goal location, but it is not in itself a representation of this goal. In rats, there is not much evidence that individual hippocampal cells encode the animal's goal location (at least, while the animal is still moving towards this goal). There appears to be some remapping of the place cells in an environment in response to changes in the location of behavioural goals (see e.g. Kobayashi *et al.*, 1997). But there are also studies which find no changes in place cell firing when reward locations are changed; see e.g. Tabuchi *et al.* (2003). The clearest single-cell evidence for goal-location-encoding hippocampal cells comes from one of the rare studies on humans; Ekstrom *et al.* (2003) found hippocampal cells which fire selectively when the agent is navigating towards a certain goal location, regardless of the agent's current location or orientation.

In any case, it is likely that representations of navigational goals also involve regions downstream of the hippocampus in both humans and other animals. Hok *et al.* (2005) have recently found place cells in the rat prelimbic frontal cortex. As in the hippocampus, a place cell is one which is active when the animal is in a particular place in its environment. These cells appear to provide a coarser representation of location than hippocampal place cells. In addition, the distribution of places encoded by these cells is strongly biased towards locations with behavioural significance. (In the experiment, there are two such locations: a 'trigger zone', which the rat has to reach in order to trigger a food reward, and a 'landing zone', where the reward arrives.) These characteristics make frontal place cells well suited for navigating to goal locations. The summed output of these cells provides a signal which increases as the animal approaches a goal. The animal simply has to follow the gradient of this signal. Interestingly, there has recently been corroborating fMRI evidence in humans that frontal areas encode a measure of distance to a goal location. Spiers and Maguire (2007) gave subjects a virtual reality navigation task, and found that activity in the medial prefrontal cortex correlated positively with their proximity to the goal. They also found activity in a right subicular/entorhinal region which was also (negatively) correlated with proximity to the goal. In sum, there is relatively good evidence for allocentric goal representations, in regions downstream of the hippocampus.

Many computational models of navigation assume a set of extrahippocampal **goal cells** which generate a gradient of activity representing distance to the goal (see e.g. Burgess *et al.*, 2000; Hok *et al.*, 2005).[17] I will assume that the agent has a map representing **goal places** in his current environment, in addition to a map representing his current place. A simple circuit supporting learning of and navigation to goals is given in Figure 13.6, roughly based on the model of Burgess *et al.* (2000). The agent's current location in the environment (or more generally, the *subject's* current location, see Section 13.5.3.1) is encoded in the current subject place units (i.e. in hippocampal place cells). Each point in the environment is also represented by a goal place (i.e. a goal cell); for simplicity, I assume a one-to-one mapping between these units. There are bidirectional links between each goal place and a global 'reward' signal. The link for each goal place is gated by the activity of the corresponding current place. Finally, I assume the agent is maintaining a

---

[17] In other models, the gradient to be followed is generated by different mechanisms. For instance, Bilkey and Clearwater (2005) hypothesise a remapping function which causes place fields to cluster around locations associated with reward, as has been found experimentally (see e.g. Hollup *et al.*, 2001). Their function generates a gradient of place cell density which can be followed to reach goal locations.

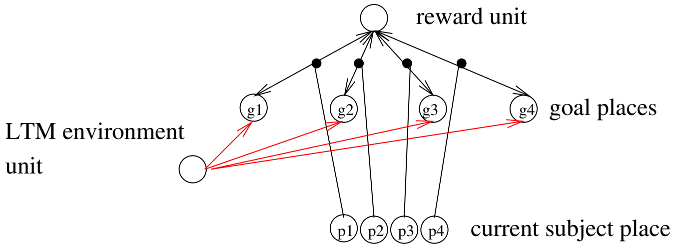


Figure 13.6: A simple network supporting learning of and navigation towards goal locations

representation of his current environment as an active LTM environment unit, which links to all the goal places (these are shown in red in the figure). If an agent is navigating the environment and happens to receive a reward, the goal place unit associated with his current place will become active. Hebbian learning will strengthen an association between the current environment unit and this goal place unit, so that it is activated whenever the agent returns to this environment. (These associations can be thought of as implementing a memory for 'desirable locations' in the current environment, analogous to the memory for object locations discussed in Section 13.4.2.3.) When exploring the environment at a later date, goal location memory will create a pattern of activity over the goal places. Activity from goal places which will be transmitted to the reward unit, via the connections enabled by active current places. The agent can then read the activity of the reward unit as an indicator of proximity to places previously associated with a reward, and can navigate in a manner which increases this activity.

A gradient following model of this kind needs to be supplemented with several additional mechanisms. I will discuss these in the remainder of this section.

13.11.3.3 Mechanisms supporting reward gradient navigation

How does the gradient of summed goal cell activity 'spread outwards' from the goal to other places in the environment, so that the agent can begin to follow it even when at a distance? There are several mechanisms which help provide a solution. One mechanism relates to the fact that current and goal locations in the environment can be represented at coarse spatial granularities as well as at fine ones. The agent can thus be quite distant from the goal, but still be 'at the right location' at a coarse granularity. As he gets closer, this coarse grained signal will be supplemented by progressively finer-grained signals. However, this mechanism is not likely to suffice for navigation over large distances within the environment.

Several theorists have suggested mechanisms by which points *close* to actual reward locations can become secondary reinforcers. These mechanisms might involve a general-purpose delayed reinforcement learning scheme, such as temporal difference learning (Sutton and Barto, 1990); see e.g. Foster *et al.* (2000). Alternatively they might involve specialised hippocampal mechanisms—for instance, the reverse replay of place cells seen by Foster and Wilson (2006) when a rat reaches a goal location (see Section 3.7.1) might have a role in generating a gradient from distant points in the environment.

Finally, there is good evidence that agents can play forward candidate trajectories in simulation when planning a locomotion action (Diba and Buzsàki, 2007; again see Section 3.7.1; see also the discussion of Johnson and Redish 2007 below, in Section 13.11.3.6). These prospective and retrospective simulated trajectory traversals may help to link the animal's current location with goal locations in the environment.

13.11.3.4 The direction-to-goal representation

A second issue for gradient-following models of navigation is that the agent has to sample the gradient in every direction at every point to discover the best direction to move in next. Several theorists propose that the agent maintains a *direction-to-goal* representation as well as a distance-to-goal representation, which obviates this problem. The strongest evidence for this in humans comes again from Spiers and Maguire's (2007) fMRI study. They found that activity in the posterior parietal cortex correlated with the angular distance of the goal from the direction currently faced by the agent. Recall that we have already hypothesised an 'orienting function' which takes an arbitrary allocentric place representation and generates an action which directs the agent's attention to the associated point in the world (see Section 13.4.2.2). Recall also that this orienting action was assumed to be implemented in parietal cortex. Spiers and Maguire suggest that direction to goal is encoded in parietal cortex as an egocentric action of attention (presumably one which can be overt or covert). In line with this idea, I will postulate an **orienting-to-goal function** which takes a *goal place* from frontal cortex and generates an egocentric parietal signal indicating the direction in which the agent must turn to get to the goal. Note that as the agent moves through the environment, his changing allocentric location and orientation will update this signal to keep it directed at the goal location.

13.11.3.5 Mechanisms for selecting a goal location

A third issue concerns how an agent navigates in the presence of multiple goals. Consider a situation where the agent is in between two places equally strongly associated with reward. If he navigates towards one, he navigates away from the other; the aggregate reward signal recorded from the population of goal place units may remain fairly flat.

Note that the problem is analogous to one discussed in Chapter 2: if an agent represents multiple reach targets as movement vectors, there must be something to prevent him reaching for a point in between the two targets. The mechanism required for reaching is competition between targets. Several models of locomotion action planning invoke a similar competitive mechanism to solve the same problem. Figure 13.7 sketches a mechanism which allows the current environment to be associated with many candidate goal places, which then compete so that a single goal place is selected to create a reward gradient for the agent to climb. (Inhibitory links between goal places are represented by lines with perpendicular ends—not all of them are shown.) Note that this model still supports the Hebbian learning of associations between LTM environment units and candidate goal units.

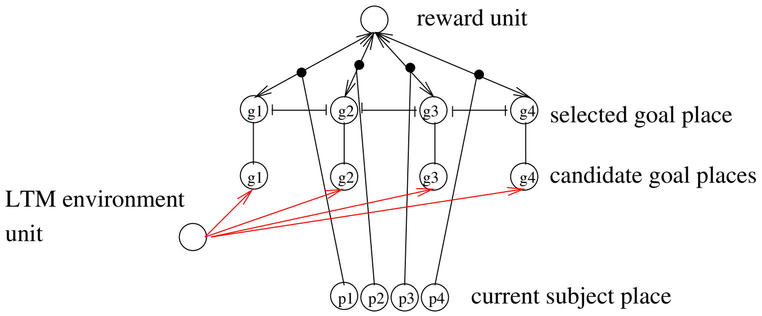


Figure 13.7: An extended network for goal places, including a competitive layer

It is useful to compare this network with the one described in Section 13.5.3.2, which describes a similar scheme for selecting a place in the environment. While the network in Section 13.5.3.2 selects a place to *attend to*, the current network selects a place to *navigate to*. Note that the current network can easily be further extended to allow an inhibition-driven serial search through candidate goal places, for instance if no trajectory can be found to a given selected goal place.[18]

13.11.3.6 Evidence for serial look-ahead search in locomotion action selection

A very interesting recent study by Johnson and Redish (2007) has found evidence that rats consider alternative target locations in sequence, and play forward a simulated trajectory to each candidate location in turn before embarking on a locomotion action. These look-ahead sequences tend to occur at junctions in a maze being run, or at points when the rat has made a navigational error—they seem to correspond to planning (or replanning) episodes.

Note that this finding sits well with a model of locomotion action planning in which alternative goal locations are considered serially, as just proposed in Section 13.11.3.5. It also sits well with a model in which the agent plays forward various trajectories in simulation to sample the reward gradient at distant locations, as suggested in Section 13.11.3.3.

13.11.3.7 Representations of goal objects

In the discussion so far, goals have been *places*—particular points in the agent's current environment. However, an agent can also have a high-level goal to reach an *object*. In this situation, the agent has to navigate to a particular location, but reaching the location is only significant because of the object it contains.

It is useful to think about high-level goals by referring back to Chapter 3. In the model of action planning described there, an agent might have an intention to make himself the agent, then establish a cup within his peripersonal space, and then to perform a reach

---

[18] I don't really get that. I think that the serial search should be of possible trajectories, rather than of candidate goals.

action. However, if there is no cup within reach, the agent might have to move around his environment to find a cup. Planning an appropriate locomotion action now depends on the agent's knowledge of the locations of objects in the room. This is encoded in object location memory, as described (most recently) in Section 13.5.3.3.

In fact, to incorporate a reference to goal locations, we have to extend the notion of object location memory still further. If we are searching for an object in the environment to *observe*, we must establish it as a subject; therefore object location memory should map LTM individuals to candidate subject places. If we are searching for an object to *reach by locomotion*, we must establish it as a *candidate goal place*. Our model of goal places already assumes a mapping from 'goal places' to 'current places'. I will propose a scheme in which the mapping from LTM individuals to places abstracts away from whether the place is a current subject place or a goal place, to allow this distinction to be contributed by a separate component, as shown in Figure 13.8. In the new scheme, object location memory is

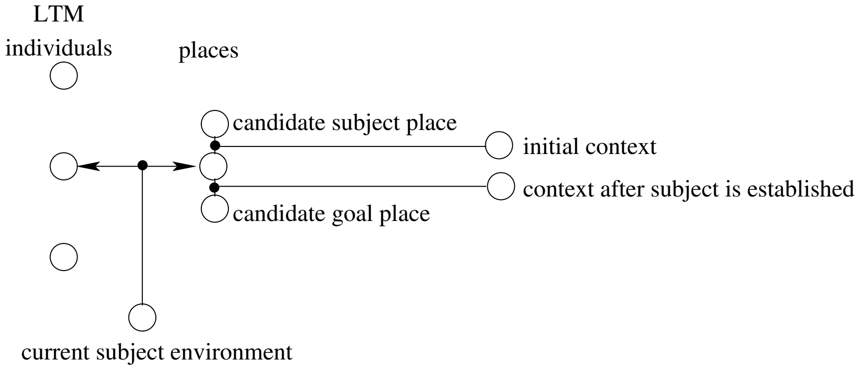


Figure 13.8: A further extended model of object location memory, including goal places

still a set of associations from LTM individuals to places, gated by the current environment representation. (Just one such mapping is shown in the figure, for clarity.) However, the mapping is now to a *neutral* representation of place, which is separately linked to candidate subject locations (as in Section 13.5.3.3) and to candidate goal locations. These links are selectively enabled by the two representations on the right of the figure, which encode how far the observer is in his current sensorimotor sequence. If he is at the very start, i.e. in the 'initial context', an active LTM individual will activate a candidate subject place. If he has already established a subject, an active LTM individual will activate a candidate goal place. This network allows several high-level goals to reach particular objects to be active in parallel, and to activate a set of candidate goal places in parallel. A single goal place can then be selected, either because of its intrinsic assoiations with reward, or indirectly, because it is remembered to contain an object associated with reward.

13.11.4 Egocentric represesentations involved in the control of locomotion actions

While allocentric representations are important in navigation, egocentric representations are also important—often more important. If an agent can see a landmark, then navigating towards it is relatively easy: its horizontal angle in the agent's visual field provides a direct signal about which way the agent should turn in order to navigate towards it. If an agent can see an obstacle, a similar principle specifies how he should turn to avoid it. There appear to be some specialised visuomotor pathways subserving online control of locomotion actions, which are quite distinct from the visual pathways involved in computing environment-centred representations.

Visuomotor routines subserving locomotion actions mainly involve the analysis of visual texture and motion energy at particular points on the retina (see e.g. Gibson, 1950; Perrone, 1992). Like locomotion actions in general, these routines can be defined in relation to the landmark being used to control navigation, and to the trajectory being pursued in relation to this landmark. In this section I will summarise the main routines.

13.11.4.1 Navigation in relation to surfaces in the environment

As discussed in Section 13.11.1, the agent can navigate using a surface in his current environment as a landmark. The surface in question might be a path through a garden, or a wall in a corridoor; most generally, it can simply be the omnipresent floor or ground. The most obvious trajectories afforded by a surface are parallel to the surface; i.e. trajectories which keep the agent and the surface at a constant distance and relative orientation. For instance, when an agent travels over the floor of a room, he travels parallel to the floor; when he travels down a corridoor, he travels parallel to the floor and to the walls.

There appear to be special visual routines which allow an agent to monitor his progress along trajectories which are parallel to surfaces.[19] These routines exploit the fact that when the agent follows such trajectories, the surface typically projects a relatively *stationary region* onto the retina. For instance, when an agent is travelling down a corridoor, the retinal regions associated with the two walls of the corridoor will each remain relatively unchanging, despite the agent's own motion.

Within the retinal region associated with a surface, there are several cues which the agent can use to control locomotion. One of these is the **texture gradient** in the region. The optic texture projected by a surface becomes finer the further it is from the agent. The gradient of texture fineness through the region can be used to compute the orientation of the surface (in relation to the agent, naturally). A second cue becomes available when the agent starts to move. Within the surface there will be a pattern of **optic flow**, which provides information about the direction of the agent's motion in relation to the surface. To take a simple case: if the surface is flat, we can imagine a notional 'horizon line' defining its limit if infinitely extended. If the centre of expansion of the agent's optic flow field is on this horizon line, he is travelling parallel to the surface. If the centre of expansion is

---

[19] References needed here.

below the horizon, he is moving towards it, and if it is above the line, he is moving away from it. Optic flow therefore allows the agent to adjust his movement to pursue a desired trajectory in relation to a nearby flat surface.

13.11.4.2 Navigation in relation to object landmarks

An agent's locomotion actions can also use landmarks which are objects situated at 'places' within his current environment. Again, there are different trajectories which can be assumed in relation to these landmarks, which can characterised by specialised visuomotor routines.

Assume an agent is navigating *towards* an object landmark. If he is moving exactly towards it, the retinal projection of the object will be at the centre of an expanding optic flow field. If the retinal projection of the object is to the right of the current centre of optic expansion, his current trajectory will leave it to his left, and he must correct his course by turning to the right; if its projection is to the left of the centre of exansion, he must turn to the right. A circuit implementing these visuomotor relationships will steer the agent to the target object. (I think here you can cite work on 'beacon homing' and 'piloting'—see e.g. Whitlock *et al.*, 2008).

If the landmark object is an obstacle, a different set of visuomotor mappings should be used. The appropriate mappings will depend on the trajectory the agent wants to take in relation to the landmark. If the agent wants to navigate *past* a given object, the object's projection should be maintained at a certain distance to the left or right from the centre of optic expansion. (The appropriate retinal distance will be a function of the object's actual distance from the agent, becoming larger as the object approaches.) If the agent wants to navigate *over* the object, its projection should be maintained at a certain distance *below* the focus of expansion, in a similar way; and if the agent wants to navigate *under* the object, its projection should be maintained at a certain distance *above* the focus of expansion.

The case of navigating *around* an object is interestingly different, since it involves describing a curved trajectory. Here the visuomotor routine involves maintaining the object's projection at a particular distance from the focus of expansion (typically to the left or right of it). However, when the object is at a certain distance from the agent, its projection should be maintained at a constant distance from the focus of expansion. This will force the agent into a curved trajectory around the landmark object. Another visuomotor routine for travelling around an object requires the agent to fixate the object, and move in a direction relative to his angle of fixation (see Land and Furneaux, 1997 for evidence that this fixation-based strategy is used by car drivers steering around corners).

Note that obstacles to be avoided can also be defined as objects which are *themselves* moving towards the agent. From the perspective of egocentric visuomotor routines, objects looming towards the agent will generate similar texture flow patterns, and the appropriate motor response can be expressed as a function of these patterns. As noted in the discussion of the reach-to-grasp action, there appear to be specialised visuomotor pathways for the avoidance of objects moving towards the head (see e.g. the discussion of the VIP-F4 pathway in Section 2.5.2.3). These routines may also be invoked when the agent needs to navigate around a stationary obstacle, at least at close quarters.

13.11.4.3 Visual attention to navigation landmarks

It is interesting to note that all of the visuomotor routines discussed above associate landmarks with well-defined *retinal regions*. Sometimes these regions are stationary in the visual field, and contain optic flow patterns. At other times the regions are moving along optic flow lines. In all cases, the routines require *reference* to a particular (moving or stationary) retinal region.

The association of navigation landmarks with retinal regions provides the basis for an interface between the attentional routines involved in navigation and those involved in object classification. Recall from Section 2.4 (and from the model in Section 10.5) that in order to classify an object, it must be the most active point in the saliency map. When we describe a locomotion action, we can identify the landmark, which presumably involves the regular process of object classification. It thus appears that when monitoring a navigation action, we can define the most active point in the saliency map as the point currently being used as the reference for our visuomotor navigation control routines. This correspondence is reminiscent of the correspondence between attention-for-motor-control and attention-for-object-classification found in the reach motor pathway, as described in Sections 2.5.2.7 and 2.5.2.8.

13.11.4.4 Trajectories in relation to landmarks in an allocentric frame of reference

The notion of a 'trajectory in relation to a landmark' seems to be quite well defined at the level of egocentric visuomotor routines, as described in the preceding sections. Is it also defined in an allocentric frame of reference? We have already seen evidence that the hippocampus holds environment-centred representations of locations by themselves (Section 13.4.2) and of trajectories by themselves (Section 13.11.3.1). We have also seen how a trajectory towards a goal location can be defined in terms of reward gradient climbing (Section 13.11.3.2) and the maintenance of a direction-to-goal representation (Section 13.11.3.4). But are there ways of representing trajectories *past* or *around* landmarks in an allocentric frame of reference? It certainly seems important for an agent to have allocentric ways of specifying such trajectories. If the agent is generating an allocentric trajectory to a goal location using only a reward gradient and a direction-to-goal representation, the generated trajectory will lead straight to the goal, ignoring any obstacles in the route. It thus seems necessary that the mechanism which generates allocentric trajectories is influenced by the location of obstacles and boundary surfaces in the environment. (Moreover, the agent needs a way of representing the trajectories followed by *other agents*, when he is watching locomotion actions, rather than experiencing them—see Section **??**.) There is recent evidence that the mapping between egocentric and allocentric representations supporting navigation is effected by posterior parietal cortex (see e.g. Whitlock *et* *al.*, 2008).

One way to encode obstacles and boundaries is to represent them as locations which contribute *negatively* to the reward gradient (see e.g. Burgess and O'Keefe, 1996; Vieville, 2006). Provided certain conditions are met, trajectories generated in an environment containing several obstacles and a single goal will reach the goal. These conditions relate to the amplitude of the 'dip' in the reward gradient generated by each obstacle, and to the configuration of obstacles in relation to the aget and goal locations. While the gradient associated with the goal locations should extend over the whole environment, the trough associated with each obstacle should be relatively local, so that it only influences the agent's navigation when he is close to it. If there are too many obstacles, it may be that there is no trajectory to the target following a monotonically increasing gradient. However, the possibility of sequential search among the available trajectories goes some way to remedying this problem.

Note that the location of obstacles cannot simply be ignored; these locations have to be represented actively, if negatively. This is again reminiscent of the case of reaching; recall from Section 2.5.2.2 that locations associated with 'distractor objects' in a reach task are actively inhibited (see Tipper *et al.*, 1998). But it is also unlikely that all the potential obstacles in an environment are represented equally prominently. Recall from Section 13.11.4.3 that visuomotor obstacle-avoidance routines involve the allocation of *attention* (in an egocentric frame of reference) to the object to be avoided. The external object location function (Section 13.4.2.1) ensures that the location of this object will receive special prominence in the agent's allocentric representation. Finally, note that reward gradients must have particular shapes for particular trajectory types. For instance, if an agent is navigating *around* a landmark object, the gradient must decrease if the agent approaches too close to it, but also if he moves too far away from it.[20]

How can an appropriate gradient pattern be learned for different trajectories defined in relation to landmark objects? One interesting possibility is that the function which generates the gradient is learned during the agent's experience, using the visuomotor routine currently controlling navigation *as a training signal*. Different routines associated with 'past the landmark', 'around the landmark' and so on will generate different trajectories in relation to the attended landmark. These trajectories will be recorded, as the agent moves from one point to another. The agent can perhaps learn a function which deforms the shape of the reward surface around his current location and that of the attended landmark so that a gradient-climbing mechanism recreates the experienced trajectory. One of the inputs to this function would be the trajectory type, defined as the visuomotor routine which creates the trajectory.

Again, note a correspondence with the model of reaching developed in Chapter 2. Recall from Section 2.7.5 that the agent develops methods for representing reach-to-grasp actions as observed trajectories of his own hand onto a target object, because these trajectories correlate with the motor representations which drive his actual movements. Once

---

[20] This gradient pattern could be generated by superimposing a a positive peak and a higher-frequency negative peak, both centred on the target object; the so-called 'Mexican hat' surface.

these perceptual representations of actions have been learned, the agent can recognise similar actions performed by other agents; these learned correspondences thus constitute the foundation for the 'mirror system' for action representation. The proposal in the current section is that allocentric trajectory representations are similarly learned as independent perceptual correlates of the sensorimotor routines which control his own actions. Once learned, these representations can then be deployed to represent the locomotion actions of external agents. However, in the current case, they also have a role in planning the agent's own locomotion actions, in situations where the agent is unsighted.

Before moving on, it is also interesting to consider the visuomotor routine of navigating parallel to a boundary surface in the environment, as described in Section 13.11.4.1. Is there a way of characterising a trajectory parallel to a boundary surface in allocentric terms? Note first that a boundary surface is not an object *in* the agent's environment—rather, it is what defines the spatial structure of this environment. One interesting possibility is that the state in which the agent is following a boundary surface is one in which the function generating activity in the agent's 'boundary cells' can be trained. The model of boundary cells touched on in Section 13.2.1.2 could possibly be expanded in this direction.

13.11.5 A model of the planning and control of locomotion actions

I have now introduced several cognitive processes involved in the planning and control of locomotion actions. In this section, I will consider how these processes interact together. As for the action of reaching-to-grasp, I suggest that executing a locomotion action involves a characteristic *temporal sequence* of processes, and accordingly a characteristic sequence of sensorimotor representations.

The agent starts off in the same neutral initial state as for the reach-to-grasp action (see Section 2.10.2), in which he must decide whether to perform an action himself, or to observe an external agent or object. In Section 13.5.3.2 I suggested that the attentional medium within which this decision is made is the set of candidate subject places—an allocentric representation of places in the agent's environment which are associated with objects, in which his own location is also represented. As for reaching, competition between candidate objects is mediated by bottom-up perceptual factors and by top-down factors, as summarised in Figure 13.5.3.2. Some perceptual stimuli encourage the agent to initiate an observing action—these mainly relate to the perceptual salience of external objects. (In the context of locomotion actions, the most important contributor to a region's salience is movement in relation to the background.) Other perceptual stimuli, such as looming objects, encourage the agent to initiate a locomotion action. Attention to an external object can also be encouraged top-down. I assume that in the initial context, there is a pattern of activation over the agent's set of LTM individuals identifying individuals which it would be worth attending to as subjects. This pattern can be channelled through object location memory to provide top-down input to any places in the environment at which activated objects have recently been seen. Top-down support for the 'attend-to-self' operation can be channelled through the current observer location function. When a single location is selected, the location is also attended to, and the object at this location can be classified. At this point there is an option for the observer to reconsider: maybe the classified object is not such a good subject after all, and the currently selected location is inhibited to look for another candidate subject. In our scenario, the observer decides to act—i.e. to establish himself as the 'current subject'—so the current observer location becomes the selected current subject place, and we enter self-observation mode.

Having entered this mode, the locations in the agent's environment change their stripes: the same locations which were hitherto competing to be selected 'as subjects' now compete to be selected *as goals*. This competition happens simultaneously in two attentional media. One is a motor representation of objects in the agent's perispace (see the account of movement vectors in Section 2.5.2). The other is an allocentric representation of goal locations, as described in Section 13.11.3.5. Some locations in an environment might be intrinsic navigation goals, because of their association with reward, as described in Section 13.11.3.2. Others might be goals because of their association via object location memory with particular LTM individuals, as shown in Figure 13.8 in Section 13.11.3.7. When an environment-based goal location is selected, a trajectory is played forward in simulation from the agent's current place, informed by a reward signal generated in frontal cortex, and a parietal representation of direction-to-goal. Again, I assume there is a serial component to the process of goal selection, as described in Section 13.11.3.6: several trajectories may be tried to a given goal location, and several goal locations can be tried in turn, so that the object at that location (if there is one) can be attended to and classified, to match to a higher-level goal representation. I also assume that particular trajectories afforded by the environment compete for selection along with goal locations, as discussed in Section 13.11.2.1. These can be thought of as 'bottom-up' contributors to the eventual selection of a trajectory to follow.

I propose that the salient regions in the agent's motor maps, depicting objects in his perispace, compete against regions in the map of the agent's environment *on an equal footing*. If the winning region is in the perispace map, the agent will execute a reach action, and if it is in the environment map, he will execute a locomotion action. In other words, the map from which the winning location is selected determines the subsequent mode of sensorimotor processing. (This idea of competition between multiple maps is similar to the proposal about competition between saliency maps in the accout of hierarchical visual attention in Section 10.8.3.3.) In our case, the winning location is in the environment-centred map, so subsequent control of the agent's movements is devolved to allocentric representations.

The final stage of the locomotion action involves the execution of the planned action. Online control of the action is provided in two frames of reference. In an allocentric frame, a hippocampal trajectory of 'current subject place' cells is traced up a reward gradient, informed by a parietal direction-to-target representation and by perturbations generated by representations of obstacles. In an egocentric frame, a visuomotor routine is engaged which allocates attention to the target or to obstacles, and generates actions contingent on particular patterns of retinal optic flow.

When the agent's navigation routines are well trained, the agent will reach the target location. (...)

Something about a standard sequence of sensorimotor representations?

13.11.6 Action execution and action observation modes and the object location function

The above discussion assumes that the agent executing the locomotion action is also the observer of the action. But what if the locomotion action is being observed by a third party? In Section **??** I proposed that the observer can exist in two modes, depending on whether he is performing an action or watching another agent. The observer's first decision when initiating a sensorimotor process is to establish one or other of these modes. A similar argument applies for locomotion actions. If the above discussion relates to the agent's own locomotions, how is the environment-centred motor controller configured for action observation mode?

In action execution mode, there is a perceptual function which delivers the agent's own location, called the **agent location function**, and a function which identifies potential locomotion landmarks in the environment, called the **landmark location function**. Both functions deliver points in the cognitive map: the former activates a single 'current place unit', and the latter activates a set of 'goal place units' (which presumably compete with one another).[21] In action observation mode, we also need a current place unit and a set of goal place units—however, the functions which deliver these units must be different.

Recall that action observation mode is triggered when an object in the environment becomes salient for some reason, and the observer decides to attend to this object rather than perform an action himself. The decision to attend to an external object in the initial context is what establishes action observation mode. I propose that in observation mode, the 'current place unit' is specified as the allocentric location of this external object, while the set of potential landmarks is given by a function which biases the 'goal place units' in the cognitive map towards those units which the agent's body is oriented towards.

13.11.7 Locomotion action perception in action observation mode

How is the agent's locomotion action tracked, or categorised, in action observation mode? For body-centred motor actions, we hypothesised a specialised biological motion recognition module to fulfil this purpose. Note that this module certainly operates when observing a third party moving through the environment. (Whole-body motions like walking and running are in fact the prototypical 'biological motion' stimuli.) However, a locomotion action is primarily defined by a chosen landmark, and a trajectory specified relative to this landmark. How are these chosen?

---

[21] Note the similarity with the representation of potential targets in the agent's peripersonal space. I should make more of that.

I propose that the inference process is similar to that involved in recognising a reach action performed by another agent. To summarise the case for recognising a reach action: the observer begins with a set of possible targets. The observed agent's gaze and body orientation restrict this set, and the initial trajectory of the agent's hand restricts it still further until only a single candidate target is left. At this point the *intended target* is recognised; the observer now monitors (a) the trajectory of the agent's hand onto the target, and (b) the agent's biological motion pattern, and these jointly serve to classify what kind of action is being performed on the target. These processes are possible because of two key invariants which link the experience of performing a reach action onesself with the experience of watching a reach action. One is a 'body movement/configuration gestalt' which can be derived either from proprioceptive dynamics or from retinal form and motion energy. The other is the fact that the hand's final movement to the target is defined in a coordinate system based on the target, which is the same regardless of whether the reach is performed by the observer or by a third party.

For the recognition of a locomotion action, I propose that there are two similar invariants. Firstly, the trajectory of the agent in relation to the locomotion landmark can in each case be explicitly generated in the cognitive map. I proposed that in action execution mode, the agent's path through the environment is explicitly computed; see Section **??**. In action recognition mode this path can also be computed, by remembering the sequence of locations which the observed agent moves through. (Recall that action execution mode and action observation mode are defined so that the agent's location is represented in the 'current place units' in both cases.) Secondly, if a locomotion action has a goal configuration, this is defined in an object-based coordinate system which is insensitive to whether the agent is the observer or a third party. (...)

13.11.8 The object location function revisited

We are assuming that the observer is able to use the cognitive map to represent the location of objects in his environment. The proposal given earlier in Section **??** was that 'object location memory' is implemented by a function which takes the currently active spatiotemporal context and returns a set of bidirectional associations between LTM individuals and cognitive map locations, so that activating an individual activates its location in that context, and vice versa. The above discussion of action execution and action recognition mode now allows us to be a little more precise.

In action observation mode, activating an LTM individual activates the associated cognitive map location *as a current place unit*. In action execution mode, activating an LTM individual activates the associated cognitive map location *as a goal place unit*. Both of these refinements have interesting implications.

Firstly, note that activations of units in the cognitive map cannot be interpreted on their own; rather, just like activations of units in the motor system, they are deictically referred to a prior action establishing either action execution mode or action observation mode. It may thus be that a given hippocampal place cell sometimes encodes the agent's current location, and sometimes the location of an object being observed. Which is encoded depends on a prior decision of the agent.

Secondly, note that the extended definition of the object location function explains how an agent can reach an object which is outside his peripersonal space. When the agent adopts action execution mode, the attended-to object will activate a goal place unit in the cognitive map, which allows the location-based motor controller to navigate to this object. Note that navigation is not driven by the object itself. The cognitive map also associates a landmark with the 'place' which the object is at, and it is this landmark which guides the agent's navigation. When the agent reaches this place, the object will be within his peripersonal space, and an ordinary reach action can be planned.[22]

13.11.9 Navigating between spatial contexts

The location-based motor controller allows the agent to plan and execute trajectories within his current spatial context. But the agent also needs a way of moving *between* contexts. I assume that there is a second, higher-level navigational controller, which plans routes which traverse contexts.

This controller requires that each spatial context the agent knows about is represented twice, once in a **current spatial context layer** and once in a **goal spatial context layer**. An active unit in the current spatial context layer denotes the agent's current spatial context. (Remember that the agent is either the observer or a third party who the agent is watching.) An active unit in the goal spatial context denotes an environment which the agent intends to get to. An active unit in the current spatial context layer can be termed the **current spatial context**, and an active unit in the goal context layer will be termed the **goal spatial context**.

I will assume that goal spatial contexts are also hippocampal context assemblies. There is good evidence that the hippocampus represents goal locations in the agent's *current* environment, as just summarised in Section **??**. I do not know of any evidence that the hippocampus also encodes goal *spatial contexts*; however, given that it encodes current contexts, and goal locations within the current context, I will assume that it also encodes goal contexts.[23]

[This all needs rewriting.]

13.11.10 Representations of extended spatial paths

It is often useful for an agent to store information about how to move between distant locations. Agents often travel some distance, particularly if a reward can be obtained by

---

[22] It is interesting to consider the linguistic correlates of this two-phase action. I expect that it can be described in two ways: either *John walked up to the cup and took it*, or simply *John took the cup*, with no reference to the locomotion action. The choice depends on whether the linguistic system is set up to interface with just the reach motor controller, or with the reach motor controller *and* the locomotion motor controller.

[23] Maybe the difference between spatial contexts and locations in the current cognitive map isn't so easy to draw. When a rat passes down one arm of a maze and a hippocampal cell fires, is this a very specific spatial context?

so doing, and can learn complex routes between distant locations. What are the representations and mechanisms which support the learning of such routes?

Recent studies suggest that areas in the basal ganglia are involved in navigating a well-learned route. Hartley *et al* (2003) present fMRI data showing that the caudate nucleus is more activated when subjects move along a well learned route than when they are navigating a new route created from general knowledge of the spatial layout of their immediate environments. This finding echoes a similar finding in rats (see e.g. Packard and McGaugh, 1996). Well-learned routes appear to be represented as stereotypical action sequences: they are established slowly, using reinforcement, and once learned, enable fast, automatic behaviour. Interestingly, while the hippocampus represents spatial environments in terms of boundaries and surfaces rather than landmarks, landmarks do seem to be involved in our procedural memory for routes. Spatial behaviour oriented to landmarks involves activation of another area of the basal ganglia, the striatum (see Doeller *et al.*, 2008), and like extended routes, is learned through reinforcement (Doeller and Burgess, 2008).

I propose that learning extended routes involves a *working memory* for sequences of spatial contexts and/or landmarks, held in the basal ganglia. I will call this **spatial context working memory**. I envisage an extra, high-level motor controller for locomotion which takes a current spatial context and a goal spatial context and delivers a sequence of 'subgoal' spatial contexts in this working memory medium. I will assume a very simple training function for the moment. In a training phase, the agent explores his environment. Whenever a new environment is entered, the associated context unit is remembered in spatial context WM. So this form of working memory holds the sequence of contexts which must be traversed to get from the first context stored to the most recent context stored. I assume that the context navigation controller is trained by the current spatial context WM. The training regime should emphasise the learning of common or behaviourally useful context sequences; this will determine when spatial context WM is reset during training, and when the current WM sequence is used as a training instance.

One interesting consequence of representing a path between distant locations as a sequence of linked intermediate spatial contexts is that representations of the relations between contexts are likely to reflect the number of intermediate links. An interesting study supporting this idea was conducted by Hirtle and Jonides (1985). Subjects were asked to enumerate landmarks in a university campus. An analysis of the sequential structure in this free recall task showed a clustering of landmarks. In a followup experiment, subjects were asked to estimate distances between landmarks. It was found that estimates were greater for landmarks from different clusters than for landmarks from the same cluster, even when the actual distances were roughly the same. This is good evidence for the kind of hierarchical structuring of contexts which would result from the indexing of one context to another.

13.12 The internal syntax of PPs

13.13 A sensorimotor interpretation of PP syntax

Chapter 14

Sensorimotor interpretations of some other clause types

14.1 Introduction

In this chapter, I will look at some of the basic clause types which have not yet been considered, and ask whether our general sensorimotor interpretation of syntax can be extended to cover these. In the first half of the chapter I will look at clause types featuring spatial PPs, drawing on the sensorimotor characterisation of spatial PPs given in Chapter 13. I will begin in Section 14.2 by considering locative clauses: stative clauses asserting that a particular object is in a particular location. (I will also include existential sentences in this category, for reasons discussed in that section.) In Section 14.3 I will consider clauses describing a simple locomotion action carried out by an agent, which feature verbs like *go*, *walk* and *run*. And in Section 14.4 I consider clauses describing more complex actions in which an agent causes a target object to move along a certain trajectory, which feature verbs like *put*, *give* and *throw*.

In the second part of the chapter, I consider some other clause types about which the sensorimotor interpretation of LF makes certain predictions, and which must therefore be discussed for the sake of completeness. Section 14.5 considers intransitive clauses, and Section 14.6 considers passive clauses.

14.2 Locative clauses, and existential sentences

[Still to be written]

14.3 Simple locomotion clauses

[Still to be written]

14.4 Ditransitive verbs and the causative alternation

(14.1) John gave the cup to Sue.

(14.2) John pushed the train along the track.

(14.3) The man opened the door.

(14.4) The door opened.

14.4.1 A sensorimotor model of ditransitive actions

There are two components to this: firstly the idea that a ditransitive action has two attentional stages which occur in strict sequence (see Section 14.4.1.1), and secondly, the idea that a transition occurs in the frame of reference used to provide feedback about the action (see Section 14.4.1.2).

14.4.1.1 The two attentional phases of a ditransitive action

Johansson *et al* (2001) study agents executing ditransitive actions which involve grasping a bar, then moving it to press a switch. In language, these actions might be described as 'moving the bar to the switch'. The actions involve two key stages: firstly a stage when the agent reaches for and grasps the bar, and secondly a stage when the bar is brought into contact with the switch. They can be described in terms of two key **contact points**: the point on the bar where the agent's hand grasps, and the point on the switch which the bar eventually touches. Johansson *et al.* analysed the eye movements of the agents, and found a strong tendency to fixate these contact points in advance of contact actually being made. Moreover, the first fixation (on the bar) is only maintained until the hand reaches the bar; very shortly after that, fixation moves to the switch. Thus at the level of attentional actions, there is quite good evidence for a strict decomposition of a transitive action into distinct stages.[1] On the other hand, the profile of the motor component of a ditransitive action seems to be specified more globally over the course of the whole action. For instance, cite that Jeannerod idea.

14.4.1.2 Proximal versus distal motor control

[2]

---

[1] However, Ballard *et al.* seem to have evidence for somewhat greater anticipation of eye movements. I should check this again.

[2] In here, you should really refer to Cisek 2005: 'The neural representation of a plan does not consist of a desired trajectory or motor program optimized prior to movement onset, but simply represents the desired motion of the controlled object in an effector-independent manner and in the reference frame of whatever defines error in the task. Most of the details of kinematics and kinetics do not emerge until after movement begins, even in very familiar tasks and even when full information about the upcoming movement is provided well ahead of time.'

Now consider a more complex manipulating action such as ‘bend’ or ‘crumple’. Any concrete action (from the perspective of the executer) ultimately involves an action involving signals to the limbs. However, the feedback which is used to learn the appropriate signal, or to regulate the signal while the action is under way, can come from different sources. For a simple transitive action like ‘grab’, it presumably comes from the agent’s proprioceptive sense of his own arm position, and from real-time visual information about the relative position of his hand and the target. But for a more complex action like *bend* or *crumple*, once the hand has reached the target, it is likely to come from the agent’s observation of the effects his action is having on the configuration of the target, rather than from any direct monitoring of his arm or even finger position.[3] If we assume there are two objects being attended to—the agent and the patient—and we assume that the biological motion system is involved in providing feedback to the motor system during action monitoring, then the suggestion amounts to the idea that the biological motion system is free to *shift* from the agent to the patient during this monitoring process. Given that the biological motion recognition system is somewhat independent of other visual systems, this seems at least possible.

14.4.2 A syntactic model using VP shells

An interesting recent account of the causative alternation is given by Heidi Harley (Harley, 2003), drawing on classic work by Larson (1988). I will quickly summarise a version of this idea given in Santorini and Kroch (2007), which can be stated quite simply.

Harley’s suggestion is that a ditransitive verb is actually a complex of a verb of causation and another verb: e.g. ‘give’ is ‘cause to have’, ‘show’ is ‘cause to see’, ‘teach’ is ‘cause to learn’ etc.[4] The idea is that the two component verbs originate at different positions in the syntax at LF; the ‘cause’ verb introduces a second VP, whose subject is the recipient argument, and whose object is the theme argument. So, the LF of a sentence like *John gave Mary a ball* should be analysed as in Figure 14.1.[5] (At PF, the subject *John* is assumed to raise out of the higher VP to the normal subject position, and the lower verb *get* is assumed to raise to adjoin to the higher verb *cause*, in which configuration it’s pronounced ’give’.) There are several attractive things about this analysis, including an explanation of similarities between causal sentences and double-object sentences in several languages. In English, it allows a nice account of the so-called causative alternation, which holds for a class of verbs like *close*, *crumple*, *bend* etc (also known as **unaccusative** verbs). These verbs can be used either transitively or intransitively:

(14.5) John bent the branch

(14.6) The branch bent

---

[3] I still don’t have a reference for this sort of ‘indirect control by monitoring of the motions of a manipulated object’.

[4] The examples are actually Santorini and Kroch’s.

[5] In Harley’s version, the second VP is actually a PP, but the basic idea is the same.

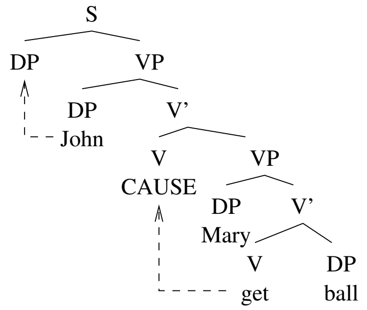


Figure 14.1: A Harley-style analysis of ditransitive verbs at LF

The alternation is odd, because the subject of the intransitive verb is the object of the transitive one. The oddness is actually with the transitive sentence; e.g. in Example 14.5 above, it's the branch that's doing the bending, even though *branch* is apparently the object of the verb. And the agent 'John' is really causing the branch to bend. Therefore the sentence can be analysed at LF along similar lines to the above ditransitives; see Figure 14.2. (The lower V still moves up to adjoin to the higher one at PF; the only difference is that (a) the combination of the two verbs is still pronounced *crumple*—so presumably *crumple* is ambiguous between 'real crumple' and 'cause to crumple'—and (b) the lower V is intransitive in this case, so there's no second object.)

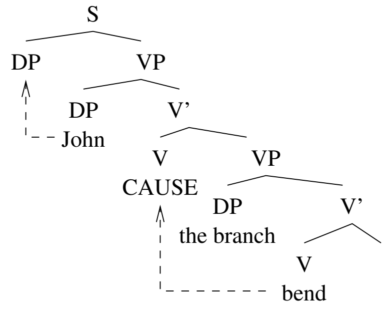


Figure 14.2: A Harley-style analysis of *bend* at LF

[6]

14.4.3 A sensorimotor interpretation of VP shells

A Harley-style postulation of two VP elements for verbs like *give* or *crumple* can be linked quite nicely to an account of proximal versus distal motor control. Here's the rough idea.

---

[6] Note: I don't think this syntactic account is compatible with Levin and Rapaport's account of unaccusatives. In their account, the branch appears as the complement of the verb at 'deep structure'.

Recall from the general sensorimotor characterisation of LF Chapter 5 that any verb phrase denotes an action which is processed in the attentional context set up by having previously attended to an agent. For an ordinary transitive verb like *grab*, the attentional context set up by attending to the agent directly sets the stage for action monitoring, and the feedback signal used to control the action relates to the body configuration of the attended-to agent. For a verb of manipulation, like *crumple*, I suggest that the agent must first execute an operation to *change* the source from which feedback for his action is derived, so that feedback comes from the motion of the (articulated) target object, rather than directly from motion of the agent's own body. So what the agent learns to do is 'whatever action results in a crumpling motion being observed in the object'. Thus, assuming a split-VP analysis of *crumple the paper* as *cause the paper crumples*, the denotation of the *cause* verb has two components: firstly an ordinary transitive action to *reach* the paper so as to be able to exert influence it (perhaps with a special requirement to grasp it in a way suited for crumpling) and secondly an operation to move the biological motion system to the paper, so that any subsequent motor action of the hand/fingers is controlled by the movements of the paper. The inner VP is then simply an observation of the fact that the paper is undergoing a certain change in configuration, described by the verb *crumple*. The fact that this observation is part of an action being carried out is captured by the syntactic context of the inner VP, which is dominated by the *cause* VP.

Note that this account also works for verbs which subcategorise for a direct object and a PP, such as *put*, understood as 'cause to go'. In this case, after having grasped the object, the agent must shift to an environment-centred frame of reference to track the object and cause it to describe an appropriate trajectory.

14.5 Intransitives

So far, we have only considered sentences describing transitive actions. Let us now look at an example of an intransitive sentence:

(14.7) The man shrugged.

Hopefully, this sentence is a little simpler to describe. The basic suggestion is that there is *no object* selected as the patient entity; rather, the motor programme can be defined solely in relation to the configuration of the agent's body. A shrug is a simple example of such an action: it just involves raising and lowering the shoulders, which can be achieved without any reference to objects external to the agent. (...)

14.6 Passives

When we were looking at transitive sentences, we could assume that the agent of the sentence appears in subject position, and the patient in object position. Of course, this is not always the case. A particularly clear counterexample is passive sentences, in which the patient appears in subject position and the agent appears only as an optional adjunct. (...)

14.7 Verbs of creation

Consider *John made a cake.* It might be thought problematic for my account, because the observer is supposed to attend to the object before starting to monitor the action—but in this case, the object doesn't exist until the action brings it into being.

At least if the observer is the agent of the 'make' action, I want to argue that the observer really does have to attend to the object in some sense before executing the action. Before you can create an object, you have to have a conception of the object *in your mind*—you can't make something unless you know beforehand what you're going to make. Clearly, 'a cake' is a *planned* object before it's a real one. I want to argue that the object exists in two modalities in this episode, just like it does in a more straightforward case like *John grabbed a cup.* If you think about it, the two modalities are even rather similar. In the case of grabbing a cup, obviously the agent attends to the actual cup to begin with. But based on this visual representation, the agent produces a representation of the cup as a *goal motor state.* This is a motor goal—a motor state that doesn't yet exist. Now think about what's involved in making an object. Let's consider a simpler case than making a cake—one that we know more about. Take *John drew a square.* We know all about how this is done, from the experiments of Averbeck *et al.* (2002). 'A square' starts off being a *motor plan.* The agent has a plan to do movement A, then movement B, then movement C, then movement D. All of the planned movements are active in parallel, and the planned square-drawing routine is selected as a single item. The point is that the square-drawing motor routine *is a square*: selecting it is like attending to a square (in the modality of motor planning).

The question now is: what does it mean to 'draw' a square? (Or to 'make' a square?) I think it means nothing more than to *execute* the selected routine. This is a nice way of thinking about the difference between drawing, making, and other verbs of creation, and ordinary transitive verbs. In a normal transitive verb, selecting the object and selecting the action are two separate things: the object contributes a single goal motor state, and there are several alternative motor programmes that are defined *with reference to* this goal state (e.g. slap, punch, grab). These motor programmes involve activation of a sequence of goal states, defined as 'perturbations' of the original goal state (see Lee-Hand and Knott, 2013). In a verb of creation, the object selected already *is* the sequentially-structured motor programme (e.g. a square-drawing plan). The 'action' done 'on' this object is just the action 'execute': quite a different sort of action.

For a verb of creation, there's no alternative way of expressing the motor planning. But for a regular transitive action there's often an interesting choice: I can say 'John grabbed a cup' or 'John made a grab [for the cup]'.[7] I can say *John moved* or *John made a movement*;

---

[7] Or alternatively, *John gave Bill a slap/a punch/a squeeze.* Clearly that's not quite the same, since you can't give someone a grab, so I'm not explaining everything about these nominalising alternations. Levin

John shrugged or *John gave a shrug.* Using the device of nominalisation, many verbs can be re-expressed in alternations featuring verbs of creation.

It's interesting to think about this in relation to the DP system. Object-creation plans in prefrontal cortex are presumably tonically active throughout the experience of the creation episode—but if they're nominalised, they can only be read out at a couple of specific times. My suggestion is that there is a cross-modal link between the prefrontal object-creation motor programme and a perceptual representation of the created object (perhaps in the form of a WM individual) and that this perceptual representation is only activated at particular points: once when the motor programme is originally selected (prior to being executed) and once ('incrementally', in the sense of Dowty, 1991) when it perceptually appears. But this idea definitely needs some work.

14.7.1 Beneficiary arguments of creation verbs: an account involving overloading

Consider *John made Mary a cake.* The beneficiary argument here (*Mary*) complicates the verb of creation. I like the idea that there are two things going on. Firstly, John is just making a cake. But he is also causing Mary to *get* a cake, or to *have* a cake. Mary has to be sentient: she must experience receiving a cake. My idea is that the causative action happens to be describable (at least in part) as making—hence it's appropriate to use the verb *make* here.

I want to say that the same thing is going on here as in *John kicked the ball into the goal.* This also involves overloading. What John is really doing is causing the ball to go into the goal. But the causative action happens to be an action that can be described as an action in its own right, as kicking.

14.8 Other thematic roles

Make some mention of Dowty's model of proto-roles, and suggest that all the properties which contribute towards proto-agenthood can be independently seen as contributing to salience. (...)

14.9 Verbs of perception

Consider *John saw Mary* and *John looked at Mary.* Here the verb describes a perceptual operation carried out by the agent on the patient (or the experiencer on the stimulus, as the participants are normally referred to). How does this fit with my model?

First assume that John himself is the observer. In the general scenario for transitive verbs, the observer first attends to himself, and then attends to Mary, and then does a transitive action (that's implicitly referred to the attended target). In this case, the

---

no doubt has more information!

perceptual action denoted by the verb ('see' or 'look at') might not seem to contribute anything new, beyond the action of attention to Mary that has already been carried out. But I'll argue that it does contribute something new. My main idea is that these verbs of perception should be thought of as describing perceptual operations. I have quite a repertoire of these, so I should be able to find ways to characterise the (sometimes subtle) differences between looking-at and seeing.

Consider *John looked at Mary* to begin with. I suggest that the action denoted by *look at* is a purely physical action, but that it's more than just the standard action of attention to Mary that's described by the AgrOP projection. The extra thing is that the physical action *endures* a little. Other similar verbs describe actions with different physical characteristics; for instance *peek at* involves looking covertly, *glance at* involves a swift look. Of course these physical actions normally come accompanied by cognitive perceptual operations. The sentence *John looked at Mary* would often be followed up with a sentence describing a state or event featuring Mary, to be interpreted from John's **perspective** (e.g. *She was happy*, *She had a brown coat on*, *She put two lumps of sugar into the cup*). I think what's conveyed here in a two-sentence discourse can also be conveyed in a single clause, sometimes with the same perceptual verb (*John looked at Mary put two lumps of sugar into the cup*), sometimes with a different one (*John saw that Mary was happy / had a brown coat on*). Note that looking at or watching an object is a schedulable operation: it can be built into a plan. It also has social connotations of its own: if I look at someone, this can communicate something to them.

Now consider *John saw Mary.* This is understood as describing John's perceptual *experience*, rather than just a physical action (see e.g. ??). My proposal is that *see* doesn't denote a particular physical action, but rather denotes the cognitive operation of going into perception mode. Normally this operation happens at the very start of a sensorimotor routine: we either go into action execution mode or action perception mode. But I suggest the operation can also 'interrupt' normal SM rocessing in action execution mode, in a somewhat non-standard way. This can be triggered by a deliberate physical action of looking (as discussed above). But it can also be triggered bottom-up: for instance if John attends to Mary in preparation for some sort of action involving her, she could *capture his attention* as a perceptual stimulus and put him into perception mode. When perception mode is selected *within* action execution mode, we get the verb *see*.

If John is someone separate from the observer, the observer establishes perception mode from the very start, and then attends to John and Mary in this mode. The observer attends to John, and establishes perception mode, and then follows John's gaze to attend to Mary, as in normal action observation. The observer may then identify John's physical action as *looking at* Mary (it's a certain sort of action); this provides a discourse opportunity to start talking about his perceptual experience (from his perspective). But the observer can also activate perception mode *within* the current SM routine. In this case I suggest the observer activates perception *in simulation* the second time (just as regular perceived motor actions are activated in simulation). But simulating perception mode is special, in that it brings about a change in the WM situation evoked by the observer. This now refpresents the observed agent (John)'s inferred *beliefs* about the situation, rather than the observer's own representation of it. There may be objects that are known to the observer but not to John; objects may be in different positions, and so on. My proposal is that when the observer registers a perceived agent's action as perceptual (i.e. activates perception mode in simulation), his WM situation changes to represent the perceived agent's beliefs.[8]

If the observer is John himself, then *John saw Mary* also involves a special activation of perception mode, but this time it's real rather than simulated. (Just like all motor actions are really executed if the observer is the agent.) Activating perception mode also triggers a change to the WM situation representation: whereas before it represented possible things John could do, now it represents the current situation more neutrally. It probably also focusses on perceptual expectations relating to Mary.

Now consider some interesting scenarios. These are partly organised by the fact that the argument of a perceptual verb can be a DP (*John saw a dog*, *John looked at the dog*), or a finite stative clause (*John saw that the dog had spots*), or a nonfinite clause (*John saw/looked at the dog bark*).

14.9.0.0.1 Perceptual verbs and false beliefs *John saw a dog* describes John apprehending the existence of a new object in the situation, in the a way that's very analogous to the existential *There was a dog.*[9] John is looking at something that's (for him) an uncategorised point in space, and classifying the stimulus there as a dog. The interesting thing is that if John is a third party being observed, *the observer* can know about the dog already: the dog is new *for John*. This is why I suggested above that activating perception mode involves establishing a new WM situation that reflects the perceived agent's beliefs, rather than the observer's own beliefs.

How can we keep track of a perceived agent's beliefs about the current situation? We know something about what they know. Say we know they have just arrived in the situation, and that they haven't been there before. Perhaps every LTM agent (or type of agent) is associated with a set of objects and locations and situations that it knows about. When an agent arrives in a situation, they don't know about it (whereas the observer may know there's a dog in it). So when they look at the location containing the dog, there's no existing WM individual representing the dog: a new one has to be created. This WM individual must somehow be tagged as belonging to the perceived agent rather than to the observer. (So later, even if the observer's own experience has updated the location stored in the dog WM individual, he can still reload John's beliefs about the situation, which may not have gone through an analogous update, and thus represent false beliefs.)

14.9.0.0.2 Nonfinite clausal complements In *John saw the dog run away*, *see* reports John's establishment of perception mode, and the complement clause directly reports

---

[8] Note that this new WM situation representation is quite similar to the situation representing the 'common ground' in a discourse (see 'New-WM/ideas'). But this contains only things which are agreed upon by both participants, so it's not exactly the same. Also, a common-ground WM situation is established by the decision to talk to an external agent, not to perceive it.

[9] This is even true of *John saw *the dog*: *the dog* has to be interpreted as a reference to a known dog, but its presence in the current situation is new.

the episode thereafter perceived. We can explain why the complement clause is nonfinite: this is because attention was directed to the dog in action execution mode, and the associated attentional action is inhibited to make room for the content of the perceived episode. We can also explain why *the dog* has accusative case.

Note we can also say *John looked at the dog run away.* This seems to put the emphasis on the physical action through which the episode was perceived. But I think it presupposes somehow that John saw the dog run away as well.

14.9.0.0.3 Finite clausal complements In *John saw that the dog was dirty*, I'm not sure that John's action of attention to the dog plays such a pivotal role. He has to attend to the dog, so that he can subsequently attend to its properties. But the dog is not the object of the *see* action.

My suggestion is that the above sentence shares the sequential structure of *John said that the dog was dirty.* In this case, John decides to talk (i.e. to enter verbal mode), then inhibits the talk episode and retrieves another proposition whose content is unrelated to the physical action. I think something similar happens in *John saw that the dog was dirty.* (i) John decides to act (i.e. establishes action execution mode). (ii) Then *straight away* he decides to observe the world (i.e. establish perception mode), before he's attended to anything. (iii) Then he perceives a state.

Why would he decide to perceive straight after having attended to himself? One idea is that facts come associated with their origin: some facts are told to me, some are witnessed by particular people. John saw that the dog was dirty: this could report a quite complex thing where John is seen to be looking at the dog, and inferred to be thinking that the dog is dirty (perhaps trivially because he says so himself). I think observing another person 'seeing *that P*' is a complex thing, involving inference of whole propositional beliefs. Consistent with this: you can't say *John looked at [the dog was dirty].* You can look at episodes—and you can look at objects (e.g. *John looked at the dog*)—but you can't look at states.

Another possibility is that seeing *that P* reports the result of an attentional operation done on an *already-established* object, hence attention to the object isn't relevant in the description and is downplayed. When I say *I saw that the dog was <u>dirty</u>*, maybe the thing I'm seeing (or better, *noticing*) is the dirtiness, not the dog. So: first I attend to myself as agent. Next I attend to the dog—but this doesn't put me into perception mode, because the dog is not salient. Now, *having attended to the dog and categorised it*, my attention is drawn to a *property* of the dog. Because this is a type of attentional capture, it's reported as the perceptual action 'see' (or 'notice'). I could say I noticed the dog's dirtiness, the dirtiness of the dog etc. I suggest I can also express the attended property within the clausal system: but in order to present the property in this system, I must present the whole predication.

I don't know if the above two ideas are alternatives to one another. I like the general idea that what you see is the informationally *new* thing in the complement clause (thus in *John saw that there was a* $[_F$ *dog] in the room* the verb of perception refers to the new information that's *added* to John's knowledge).

One interesting thing is that in *John saw that P*, *P* has to be aspectually stative: if I say *John saw that Mary grabbed a cup*, the grabbing episode is reported as a fact, rather than as something that updates the reference time. On the other hand, *John saw Mary grab a cup* is aspectually an event: the seeing happens simultaneously with the grabbing, and the two episodes move us forward in time.

Chapter 15

Adverbs and other modifiers

Some key references are Alexiadou (1997); Cinque (1999); Ernst (2002). A good summary is given in Edelstein (2012).

15.1 Manner adverbs

Here are two examples:

(15.1) John grabbed the cup angrily / John angrily grabbed the cup / Angrily, John grabbed the cup.

(15.2) John grabbed the cup quickly / John quickly grabbed the cup / Quickly, John grabbed the cup.

Manner adverbs indicate a *state* of the agent, that is *exhibited* during an action, even though the action has some other goal as its primary goal. (I think Geuder, 2002 is a good reference for this idea.) An agent doing some action can be in different states, and these will make the action come out differently.

In terms of motor implementation, I suggest that the motor system has general parameters that dictate how fast, jerky, violent etc movements will be. These are defined independently of individual motor programmes. In my model, motor programmes are defined as perturbations of goal motor states associated with targets (and perhaps beneficiaries or goal locations). I suggest that the general parameters relate to the manner in which these goal motor states are achieved: they can be achieved slowly, gently, violently, roughly, angrily etc.

At the same time, actions provide a method whereby an agent can demonstrate an internal state, whether volitionally or nonvolitionally. It's not just about the movements made; there can be things that accompany the movement, in particular facial expressions, grunts, other body language, that convey emotions or other states.

My main idea is that an agent doing an action in some manner conveys the manner and the action in parallel: the manner is present from the outset, and continues to be expressed throughout the action. In my scheme, the predication system and the action system are somewhat independent: I think the predication system identifies the manner, while the action recognition system identifies the action (and abstracts away from the manner). The predication system works using movements: it picks up on certain physical properties of the movements (jerkiness, smoothness, slowness etc), and abstracts away from the category of the actions executed. At the same time it picks up on facial expressions, which could be static.

Something that's consistent with this idea is that adverbs derive productively from adjectives (*quickly* is *quick+-ly*). Saying *John shouted angrily* is very like saying *John shouted* (the action part) and *John was angry* (the predicative part). So: where does the adverbial suffix *-ly* come from? This must indicate that the predicate was identified dynamically, by analysing the pattern of motion constituting the action.

Why can adverbs appear where they do? E.g. *John slowly grabbed the cup*, *John grabbed the cup slowly*, *Slowly, John grabbed the cup.* Note the position above VP but below the subject is interesting, because there's variation between languages: in French we have *John grabbed slowly the cup.* So there are three basic positions. I want to say these positions are positions that make sense both within the motor-action system and the predicative system.

15.2 Adverbs expressing properties of the event

Adverbs like *luckily*, *mercifully obviously*, etc predicate a property of the perceived episode as a whole (is lucky, merciful, obvious) rather than of the agent specifically.[1] (Note this can also be done in a pure predication: *It is lucky/merciful/obvious/etc that P.*) These adverbs can appear in the same positions as manner adverbials, I think. This should help us think about the nature of the interface between the predicative system and the event representation system. Whatever it is, it underspecifies whether the predicate applies to the agent/subject of the episode or to the episode as a whole.

Actually, even adverbs like *quickly* or *angrily* can be understood as applying to events. For instance, from *John grabbed the cup quickly* we can infer that *John* was quick, but also that *the grab* was quick (see again Geuder, 2002). If the nominalisation *the grab* is understood as referring to an event, then the adjectival predicate also applies to events.

Geuder notes that if someone is slow, it's because their *actions* are slow—so he argues the property 'slow' applies most directly to events (Davidsonian events), and only through some kind of metonymy to individuals.

Geuder also notes that *A beautiful dancer* is ambiguous: it could mean a dancer who is beautiful, or someone who dances beautifully. I agree—but I think if you dance beautifully, then you are beautiful *while you're dancing*: it's not just the dancing that's beautiful.

---

[1] According to Jackendoff (1972), the property of the episode is also related to the *speaker*, rather than to any participant, but I agre with Geuder (2002) that this isn't necessarily the case: it could relate to the protagonist, who needn't be the speaker.

15.3 Agentive adverbs

Consider *John kindly grabbed the cup.* When you do something *kindly*, does your manner of doing it indicate the kindness? Not always: sometimes what's kind is just the fact of your doing the action. Here from the perspective of perception the predicate *is kind* is derived by inference from the action: it's only inferred after the action is understood. (Even then, he's not necessarily *always* kind: we're not inferring an individual-level predicate. We're inferring that through doing the action he's *being kind.*) This type of adverbial is called **agentive**. The fact that the adverb can still appear high suggests that what's happening is that a stative predicate is being assembled along with an action/event representation, and then the combined structure is read out.

A purer case of an agentive adverbial is *stupidly.* In *John stupidly grabbed the cup* it's hard to see how the adverb can ever indicate the manner of the action: rather, we are evaluating John's action of grabbing the cup as being stupid. (Again we're not attributing an individual-level property to John: we're predicating a property of his action.) This is Geuder's (2002) definition of agentive.

Pylkkänen's (2002) examples of agentive adverbs are *on purpose* and *willingly.* These are again different: you can't say *It was willing of John to grab the cup*, but you can say *It was stupid of John to grab the cup. Accidentally* would be another example of this class of adverb.

15.4 Resultative adverbs

They loaded the cart heavily means they caused the cart to become heavy.

15.5 Adverb ordering principles

Event/proposition-level adverbs have to appear higher than manner adverbials.

(15.3) Obviously, John quickly left the room.

(15.4) Quickly, John obviously left the room.

Cinque (1999) suggests an elaborate structure of functional projections to explain these ordering constraints. But in my model, where each XP has to describe a SM (or at least cognitive) operation, I don't have the same latitude to postulate XPs. So I have to posit an alternative explanation.

15.6 Instrument modifiers

A VP can be modified with a PP indicating the instrument being used:

(15.5) John grabbed the cup *with the tongs.*

(15.6) With the tongs, John grabbed the cup.

The PP somehow identifies the motor system that's being used. The tongs are part of a hand motor system.

15.7 Time/place adverbs

I think these are quite different: they simply establish the temporal or spatial context for an episode. This relates to how it's saved (or retrieved) from long-term memory.

Chapter 16

Nested clauses

I want there to be a chapter describing different kinds of nested clause.

16.1 Infinitive clausal complements

16.2 Finite clausal complements

These are introduced by mental state verbs like *believe*.

16.2.1 Neural representations of beliefs

A lot of this is in the medial PFC (see the review by Wagner *et al.*, 2012).

16.3 Relative clauses

Chapter 17

A model of situations and discourse contexts

In the model presented so far, there are mappings from LTM objects to their WM and sensorimotor counterparts—e.g. LTM individuals and WM individuals, LTM environments and the observer's current spatial state. However, we do not yet have any sensorimotor or working memory correlate for individual *temporal contexts* in LTM. Recall from Section 13.9.1 that an **individual situation** is an LTM assembly which represents a particular point in time; a spatiotemporal context is formed from a LTM environment (which can be revisited at different times) and an individual situation (which can only be revisited by entering memory mode). It is interesting to consider whether there is a working memory correlate of individual situations.

Basic idea: PFC holds WM representations of candidate plans, from which the observer's winning plan is selected (see Section **??**), and it also holds WM representations of

17.1 Introduction

The aim of this chapter is to integrate the sensorimotor model of reach-to-grasp events presented in Chapters 2–3 with the model of attentional actions presented in Chapter **??**, and to use the resulting model to give a more complete sensorimotor interpretation of LF structure.

Section 17.2 presents a schematic version of the sensorimotor model, expressing the sensorimotor system as a number of modalities, each of which delivers a characteristic *representation*, generated by a characteristic *function*. Section 17.3 summarises the ideas about memory representations which we have introduced so far. One of the key ideas is that there are distinct memory systems for *individuals* and for *episodes*: there are separate working-memory media for individuals and episodes, and there are likewise separate long-term memory representations of individuals and episodes. Section 17.7 presents a model of how the sensorimotor and memory representations connect, for individuals. Section 17.8

does the same for episodes. It introduces the idea of a **situation** and a **situation type**, and the **context update function**. Section **??** considers the relationship between WM individuals and WM episodes. Section 17.10 looks at hierarchy in situation structures, and its relationship to certain structures in narrative discourse. Section 17.11 looks at how the structure of WM situations can function as a very general reinforcement learning architecture. Finally, Section **??** presents a more thorough sensorimotor interpretation of LF structures, which draws on all of the above.

17.2 A schematic sensorimotor model

In this section I will re-express the sensorimotor model presented in Chapter 2 and Chapter **??**, adopting a more formal style of description, and referring only obliquely to the neural implementation. The goal is mainly to introduce some new terminology for referring to the sensorimotor mechanisms already introduced, and to propose some generalisations about these mechanisms.

I will characterise the sensorimotor syetem as a collection of related **neural modalities**. A modality can be thought of in three ways. First, it is a reasonably coherent neural region. Second, it holds a characteristic kind of *representation*. Third, it implements a characteristic *function*, which delivers this representation, using inputs which can come from the senses or from other modalities. Thus one modality can depend on another. Clearly, the functions associated with the different modalities have to be learned, so that the representations produced are well coordinated. One of my interests is in looking at the constraints which allow this kind of coordination to be learned.

Modalities are divided into three different types, which I will consider in turn. Section 17.2.1 considers modalities involved in the representation of scenes and of the location of objects. Section 17.2.2 considers modalities involved in attending to and categorising objects. Sections 17.2.3 and 17.2.4 consider more derived modalities involved in representing events and states involving, or relating, individual objects.

17.2.1 Location-based sensorimotor modalities

I assume four special-purpose modalities involved in establishing the location of objects.

- The **saliency map** is a modality representing salient regions in the agent's visual field. I assume it is found in LIP and FEF. I assume a **saliency map function** which maps the retinal input onto the saliency map. I also assume a winner-take-all operation which identifies the **most salient region**, and an **inhibition-of-return** operation which moves to the next-most salient region.

- The **object-centred location** modality provides a representation of the space around an object, defined in a coordinate system centred on that object. As discussed in Section **??**, evidence for this area comes from studies on object-centred neglect (see e.g. Driver, 1999) as well as single-cell studies in monkeys (see especially Chafee *et* *al.*, 2007); the consensus is that representations centred on an attended-to object are developed in posterior parietal cortex.

- The **scene representation** modality provides an egocentric representation of the agent's current local environment. I assume this provides a representation which allows the agent to retrieve a stored **spatial context** (i.e. recognise the place he is in—see Section 17.3.3.2) and also to generate (and update) a representation of his own location within this context, in the **cognitive map** (see immediately below). I assume a **scene representation function**, which takes input from the whole retina, without any attentional modulation, and delivers a scene representation.

- The **cognitive map** modality provides an allocentric representation of the location of objects in the agent's current environment. I assume an **agent location function**, which takes input from (i) the scene representation modality; (ii) dead reckoning and (iii) the currently active spatial context and delivers a point in the cognitive map representing the agent's current location within the environment. I also assume an **orienting function**, which takes a region in the cognitive map, plus the current spatial context, and generates an **orienting action** establishing the corresponding area in the environment within the current visual field, as the most salient region. A minimal orienting act might just involve the selection of one of the regions in the existing saliency map. However, an orienting action can also involve large movements, such as rotation of the head or whole body, and perhaps also movement through the environment. At the end of a (successful) orienting action, a correspondence has been brought about, so that the currently dominant region of the (egocentric) saliency map is the same as the currently active region in the (allocentric) cognitive map.

17.2.2 Object-based sensorimotor modalities

I envisage three object-based modalities.

- The **cardinality** modality represents the spatial frequency currently being allowed into the object classification system in IT. I assume it is found in the intraparietal sulcus (c.f. Nieder and Miller, 2004). I assume a **cardinality function** which takes input from the saliency map (specifically, from the current most salient region), and delivers a spatial frequency—either the default one for categorising the region as a whole (which I will term the 'singular' frequency) or a higher frequency suited to categorising its texture elements, assuming they are homogeneous (which I will term the 'plural' frequency). I also assume an **inhibit current spatial frequency** operation, which shifts from the default spatial frequency to the higher frequency.

- The **object category** modality represents the results of categorising the retinal stimuli in the current most-salient region, at the current spatial frequency. I assume it is delivered in IT, and in earlier visual areas (which might deliver certain specialised properties such as colour or size). I assume an **object categorisation function**, which takes input from (i) the retina; (ii) the saliency map (the most salient region), and (iii) the cardinality modality, and delivers a set of category representations. I also assume a winner-take-all operation which delivers the **most active object category**, and an **inhibit-current-category** operation which activates the properties of the attended-to object which most distinguish it from a prototypical instance of its category.[1]

- The **object shape** modality computes a 3D representation of the shape of an object, which is used for computing grasp affordances (see Section 2.5.3). It is computed in the caudal intraparietal sulcus and AIP.[2]

I also assume top-down actions of attention in each of these modalities, which can influence the representations which are established.

- A **top-down saliency bias** can provide an additional input to the saliency map function.
- A **top-down cardinality bias** can provide an additional input to the cardinality function.
- An **top-down object category bias** can provide an additional input to the object category function.

During sensory perception, these biases can only influence competition between representations which already have some bottom-up activation, so that no illusions are seen. However, during simulation mode, these actions are sufficient to create representations in their respective modalities.

17.2.3 Event-based sensorimotor modalities

I envisage two event-based sensorimotor modalities.

- The **mirror mode** modality relates to how the agent's mirror system is configured. There are two possible configurations: action recognition mode and action execution mode. I assume a **mode-setting function** which takes input from the retina, and establishes one of these configurations. I assume the neural circuitry which implements this function involves the anterior insula and the inferior parietal cortex (see Section 2.8).
- A **motor schema** modality represents actions (either executed, observed or simulated). I assume these representations are generated in premotor cortex. I assume a **motor schema function**, which takes input from (i) the retina; and (ii) all the object-based modalities, and produces a motor schema as output. The object-based modalities provide information about one or possibly two prior actions of attention to objects (the agent and possibly patient of the action being categorised).

---

[1] I need to refer to Grossberg's ART model as the precedent for this idea.

[2] This modality doesn't play such a role in the high-level model—it could be left out.

Again, I assume that both the mirror-mode function and motor schema function can receive top-down biases.

- A **top-down mirror mode bias** influences the decision about which mode to go into.
- A **top-down motor schema bias** influences the decision about which motor schema to activate.

The mode bias can establish a mirror mode by itself, without any bottom-up input. In action execution mode, the motor schema bias can also operate by itself, without bottom-up input. But in action perception mode, it requires bottom-up input, to prevent the generation of illusions. Both biases can operate by themselves in simulation mode.

17.2.4 Attentional operations

Finally, I assume a battery of special **attentional operations**, which provide various special ways of attending to objects. The 'default' mechanism for attending to an object is via the inhibition-of-return operation, which initiates a completely separate action of attention to another object in the current scene. However, there are several methods which allow us to attend to an object in a way which is sensitive to our current attentional state. These allow the establishment of stative relationships between objects, or of stative properties of objects. The first three are named after the English closed-class words which denote them.

- The **BE** operation maintains the currently active region in the saliency map, but executes the inhibit-current-category operation. The result of this is a new categorisation operation, which establishes the properties of the currently attended object which most distinguish it from the prototypical object of its type.
- The **HAVE** operation maintains the currently active region in the saliency map, and establishes a new 'mini saliency map' bounded by this region, using a finer spatial frequency. For instance, if the agent is currently attending to a table, the HAVE operation establishes a new saliency map in the retinal region occupied by the table, in which objects associated with the table compete for attention. The HAVE operation can also create biases within the object classification system, encoding expectations about the orientation of objects in the region based on knowledge about stable contact or support relations. For instance, a cup resting on the surface of the table is most likely to be sitting on its base, and very unlikely to be balancing on the tip of its handle. Recall from Section **??** that IT probably represents object categories using sets of orientation-specific templates. Assume that the set of templates in IT are organised by orientation as well as by category. We can then envisage a function called the **pose expectation function**, which takes a region of the 'table-shaped saliency map' and applies a bias in IT towards templates with the expected orientation(s). In

summary: the HAVE function creates a mini saliency map, and imposes expectations about the orientation of objects found within this map.

- The **OF** operation maintains the currently active region in the saliency map, and executes the 'inhibit current spatial frequency' operation (described above). Its function is to allow an agent attending to a single object to categorise the group of texture elements of which this object is composed. It will only succeed if these elements are homogeneous in type.

- There is also a set of primitive **configurational** attentional operations, which are associated with locative prepositions like *in*, *on*, *under*, *by* and so on. These operations are defined in relation to an existing attentional context—they assume that the agent is currently attending to an object—and they yield locations defined relative to this object. Like the pose expectation function, they can also deliver biases to the object classification function, which encode expectations about the orientation of objects found in the associated location.[3]

17.3 Summary of the WM and LTM models

The above sensorimotor operations allow an agent to construct memory representations of the world. In Chapters 3 and **??**, a model of these memory representations was presented. A key feature of this model is that there are separate memory systems for holding memories of individuals and of episodes. This distinction holds both for working memory and for long-term memory. In this section I will summarise the memory model established so far. In Section 17.3.2, I will consider the model of memory for individuals, and in Section 17.3.3, I will consider the model of memory for episodes.

---

[3] Old material: the **location-establishing** operation is associated with locative prepositions like *in*, *on*, *under*, *by* and so on. It takes a currently attended-to object (which I called the **indexed object**) and finds its spatial relationship with a second object (which I called the **indexing object**). One of the main functions of this operation is to represent the location of the indexed object in relation to another known element of the scene, to facilitate reattention to the object later, if circumstances demand this. The initial attention to the object might be simply due to bottom-up salience, which does not in itself require any relationship to be computed. I will assume that salience-driven actions of attention are typically accompanied by a location-establishing operation.

The location-establishing operation involves three steps. First, the indexing object is attended to. Second, the HAVE operation is executed, which establishes this object as a saliency map and generates expectations aobut the orientation of objects found within this map (see above). Third, the indexed object is *re-established* as one of the items in this local saliency map. The position of the indexed object in the local saliency map, and the orientation at which it is established, determine the spatial relationship used to link the indexed object to the indexing one. For instance, *on* denotes that the indexed object is above the indexing object, and has an orientation suggesting that the indexing object is supporting it; *under* denotes that the indexed object is below the indexing object.

The location-establishing operation might involve an element of sequential search before it is successful, since it involves judicious selection of an appropriate indexing object.

17.3.1 Locations in WM and LTM

I will first summarise the model of WM and LTM for locations, which was introduced in Section **??**.

17.3.1.1 Characterisation of LTM locations

Each current LTM location is defined in two ways. Firstly, it is associated directly with one or more patterns of activation in the scene representation modality (see Section 17.2.1). These links are what enable the agent to perceptually recognise his current location. (There are several patterns of activation, enabling recognition to occur from different positions and orientations in the environment.) Secondly, it is associated with a number of goal LTM locations, to which it is **linked**. Current location LTM1 is linked to goal location LTM2 if there is an attentional action the agent can perform when in LTM1 which will lead to establishment of LTM2 (i.e. which will lead to LTM2 becoming the current location). For instance, consider two environments, Room1 and Room2, connected by a doorway. Room1 is associated with a pair of LTM locations, a current location LTMC1 and a goal location LTMG1; Room2 is similarly associated with a pair of locations LTMC2 and LTMG2. is in Room1, his LTM1 unit will be active, and the doorway will be represented as a point in his cognitive map. Firing this cognitive map location as a goal location will steer the agent to the doorway. When he reaches the doorway, his scene perception modality will establish a representation which registers his presence in Room2, and will directly activate his LTM2 location.

The linking relations between LTM locations are set up so that when a current LTM location is active, it enables a set of bidirectional associations between pairs of goal LTM locations and cognitive map locations, as shown in Figure 17.1.

`location-linking.pdf`

Figure 17.1: Linking relations between LTM locations

17.3.1.2 WM for locations

At any time, the agent must have an active representation of his current location.

17.3.2 Individuals in WM and LTM

In Section 3.6.4, I presented an account of how individuals are represented in long-term memory, and in Chapter **??** I considered how they are represented in working memory. I will briefly summarise these accounts.

17.3.2.1 Working memory for individuals

The model of working memory for individuals introduced in Section **??** was founded on the idea that perceptual establishment of an individual requires a sequence of sensorimotor operations, just as experience of the cup-grabbing event does. I proposed that when an object is established, the attentional sequence used to establish it is stored as a sequence plan in a special form of working memory called **WM for individuals**. The sequence involves three steps: establishment of a location, establishment of a cardinality, and establishment of a category.

To take an example: imagine that an agent has just established a group of dogs in a certain location L1. The associated WM individual will be an attentional plan, to establish a particular location, then to establish cardinality 'plural', then to establish a bias towards the object type 'dog'.

As outlined in Section **??**, the location of an individual can in fact be represented in several different modalities, which range from egocentric to allocentric. At the egocentric extreme of the spectrum, the location of an individual can be specified as a point in the saliency map. Moving in the allocentric direction, an object's location can be specified in relation to another object, using a coordinate system centred on that object. At the allocentric extreme, object location can be specified in an environment-centred coordinate system, as a point in the agent's cognitive map. I assume that these location mechanisms operate in parallel, and can all be associated with a WM individual. The situation is thus as shown in Figure 17.2.

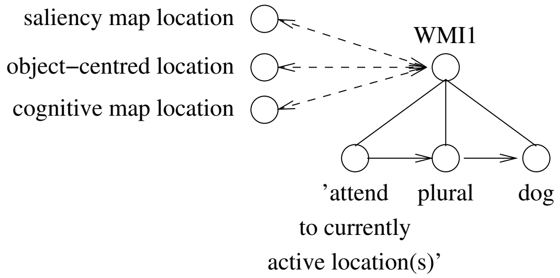


Figure 17.2: Parallel association of a WM individual with egocentric and allocentric location representations

17.3.2.2 LTM for individuals

Individual objects are also represented in long-term memory; it is for this reason that we can recognise an object currently being perceived as one which we have seen before. In the model I outlined in Section 3.6.4, LTM representations of individuals are assemblies of cells in the perirhinal cortex (see Eichenbaum *et al.*, 2007; Diana *et al.*, 2007).

As described in Section **??**, each LTM individual is associated on one hand with a WM individual, and on the other hand with a set of individual spatiotemporal context representations (as defined in Section 3.6.2). It is through the WM individual that the LTM individual is associated with sensory properties, to do with location, cardinality and type. However, these are not enough to define an individual; there could be several different objects (or groups of objects) which look identical, and thus are indistinguishable on these terms. The context representations are what keep these individuals distinct. Say I represent my own kitchen at the present time with spatiotemporal context $C1$ and my friend's kitchen at the present time with context $C2$. If there are identical red cups in the two kitchens, they will both be associated with the same WM individual; nonetheless I must create two separate LTM individuals $cup1$ and $cup2$ to represent the two cups. The LTM individuals can be distinguished by linking them to different contexts: $cup1$'s associations will include $C1$, and $cup2$'s associations will include $C2$.

An individual object is **recognised** if a combination of context and perceptual representations activate a LTM individual.

To do this, I can create an association between red/cup, $C1$ and $cup1$, and another association between red/cup, $C2$ and $cup2$. This way, the red/cup sensory representation will activate $cup1$ in context $C1$ and $cup2$ in context $C2$. In either case, the triggering of an LTM individual by a sensory representation means that the object before the senses has been recognised.

The model of LTM for individuals also included a model of memory for object locations. While the location of individuals in working memory gives emphasis to egocentric representations (in particular the saliency map), the location of individuals in LTM is primarily allocentric. Two allocentric forms of location LTM were discussed. One form stores the location of an object in a cognitive map. As discussed in Section 3.6.4.2, this type of memory takes a spatiotemporal context, and maps an LTM object onto a region in the cognitive map, or a region in the cognitive map onto an individual object. We can refer to this form of long-term memory as **environment-centred object location memory**. Another form of LTM for location encodes the position of one LTM individual in relation to the coordinate system of another object. This form of location memory can be termed **object configuration memory**. As discussed in Section **??**, whenever there is a shift in spatial attention from one object to another, it can be classified as one of the 'configurational' attentional operations: 'in', 'on', 'under', 'by', and so on. Each shift is stored in object configuration memory. This form of memory can be characterised as a function which takes an individual spatiotemporal context, *plus an LTM individual* representing an object in that context, and returns a mapping from LTM individuals to configurational attentional operations. Thus if an agent is in a certain room on a certain day (a context represented by $C1$), and is looking at a table (represented by LTM individual $Table1$) and has his attention drawn to a cup resting on the table (represented by LTM individual $Cup1$), the attentional shift involved would be classified as being of type 'on', and his object configuration memory would learn that when attending to $Table1$ in context $C1$, the 'on' attentional operation will yield $Cup1$, and activating $Cup1$ top-down will activate the 'on' operation.[4]

---

[4] In the section where you introduce this idea, you should include discussion of how you can conduct an

Note that both forms of LTM for object location this memory is useful in guiding our current behaviour as well as in recalling information from the past. If we know what spatial context we are in, and we are in a relatively unchanging environment, then object location memory tells us about the location of objects in the current scene. If we activate a particular individual object representation, then the object location function will specify where in our current environment this object is (so we can re-attend to it—or more precisely, **orient** to it, in the sense described in Section **??**). Conversely, if we attend to a particular location in the environment, the function will tell us what object to expect at this location. Similarly, object configuration memory tells us about the location of objects indexed to a currently-attended object. If we know what spatial context we are in, and what object we are currently attending to, then object configuration memory will generate expectations about the objects we will establish if we execute a particular attentional operation (like 'on' or 'under'), or conversely, what attentional operations we have to execute if we want to establish a particular goal individual.

17.3.3 Episode representations in WM and LTM

In Chapter 2, I proposed that an agent experiences a cup-grabbing event as a sequence. In Chapter 3 I suggested that our working memory of a cup-grabbing event also takes the form of a sensorimotor sequence, as too does our long-term representation of the event. The working memory medium in which episode-representing sequences are stored was termed the **episodic buffer**; it was assumed to involve an area in prefrontal cortex. The medium in which long-term episode-representing sequences are stored was termed **episodic memory**; it was assumed to involve the hippocampal system and the temporal cortices. In this section I will summarise the way the cup-grabbing episode is represented in working memory and long-term memory.

17.3.3.1 WM episodes

The WM representation of the cup-grabbing episode is a planned sequence of sensorimotor operations. The first two operations are actions of attention, to the agent and the patient respectively. Recall from Section 17.3.2.1 that each action of attention is itself stored in WM-for-individuals as a planned attentional sequence. So the first operation in the episode-level sensorimotor plan is *the activation of an attentional sequence plan*, and the second operation is the activation of another attentional sequence plan. What we have here is a hierarchical structure of action plans—a concept which is well established in models of motor control. I will discuss this hierarchical structure more in Section **??**.

---

internal search for objects by stepping through object configuration memory. Obviously you should also say that it forms the basis of memory of locative statements like *The cup is on the table.*

17.3.3.2 The structure of an LTM episode: context representations and LTM individuals

In the model of episodic memory, there are two key ideas. One is that episodes in memory are individuated by being linked to specific **spatiotemporal contexts** (see Section 3.6.2). The other is that episodes in memory are stored ‘as sequences’ (see Section 3.7).

What does it mean to say that the cup-grabbing episode is stored ‘as a sequence’? It is more accurate to say that the episode is stored as a set of weights in the hippocampus, which ensure that when an appropriate cue is presented, a certain sequence of representations is generated. For the cup-grabbing episode, the generated representations are, in turn, a LTM individual representing the man, a LTM individual representing the cup, and a premotor cortex assembly representing ‘grab’.

What does it mean to say that the cup-grabbing episode is linked to a particular spatiotemporal context? Recall from Section **??** that each episode-denoting sequence is associated with a ‘macro-context’, which when activated functions as a cue to replay the sequence. The last item in the replayed sequence, together with the current macro-context, triggers an update of the macro-context, to represent the spatiotemporal context resulting from the completed event. A completed event thus moves us into a new spatiotemporal context.

17.3.4 Summary

To review thus far: there are four different types of memory which we can refer to in our sensorimotor model: **WM-for-episodes** (also known as the **episodic buffer**), **WM-for-individuals** (also known as the medium in which **object files** are stored), **LTM-for-episodes** (also known as **episodic memory**) and **LTM-for-individuals** (also known as **recognition memory**). These different forms of memory are summarised in Table 17.1.

WM-for-individuals		**WM-for-episodes**	
also called:	‘object-file memory’	also called:	‘the episodic buffer’
form:	planned SM sequences	form:	planned SM sequences
location:	PFC(i)	location:	PFC(ii)
LTM-for-individuals		**LTM-for-episodes**	
also called:	‘recognition memory’	also called:	‘episodic memory’
atomic units:	LTM individuals	atomic units:	individual spatial contexts individual temporal contexts
relates:	SM representations	relates:	contexts, LTM individuals, actions
location:	PRc	location:	hippocampus/PHc

Table 17.1: The four types of memory in the sensorimotor model

17.4 A function-based characterisation of the agent

In this section I will formally specify the operations carried out by the observer's cognitive system in several of the example scenarios we have considered, making reference to representations in the sensorimotor system and in the WM and LTM systems. In Section 17.4.1 I introduce some top-level functions for establishing the observer's current environment, location and situation. In Section 17.4.2 I consider functions involved in categorising and individuating objects. In Section **??** I consider reaching-to-grasp, and in Section 17.4.4 I consider agent locomotion.

17.4.1 Top-level functions

The **allocentric observer state recognition function** takes the representation delivered by the **scene representation** perceptual modality[5] and computes the **allocentric observer state**, which comprises the **allocentric observer location** and **allocentric observer orientation** (both expressed in a coordinate system centred on the scene). Likewise, the **allocentric external object state recognition function** takes the scene representation modality, the allocentric observer state, and the observer's current **body-centred gaze direction**, and returns the location in the cognitive map associated with the object or place being attended to.

The **environment recognition** function takes retinal input and a representation of the observer's **most recent LTM environment**, and returns a representation of the observer's **current LTM environment**. (See Section 13.2 for details of this function.)

The **WM situation recognition function** establishes the observer's WM situation. It's normally an update from the previous WM situation (see Section 17.8.1.2). But in some cases there's discontinuity, and it's computed from basic drives (e.g. for food or shelter), and from the current environment type (certain environments automatically put you in certain situations).

17.4.2 Cognitive processes involved in object individuation

I assume that these processes function whenever focal visual attention is established, which can happens several times in a cognitive routine.

The **WM individual matching** function takes the current most salient visual region and activates one or more WM individuals which are currently associated with that region. This association may be direct, or indirect, via a FINST tracking process.

The **cardinality establishing** function takes the current most salient visual region and a top-down **cardinality bias** from active WM individuals, and returns a **categorisation spatial frequency**—i.e. the frequency of the primitive visual features which are used as input to the object categorisation function.

---

[5] It probably also takes a *working memory* representation of the observer's current speed, plus a representation of the observer's *locomotion motor command* at the previous time point, to contribute a path-integration component.

The **object categorisation** function takes the current most salient visual region, a categorisation spatial frequency and a top-down **category bias** from active WM individuals, and returns a **current object category complex**. The function does not operate until a categorisation spatial frequency has been assigned.

An **object individuation** function takes the current object category, the current LTM environment, and the current location within this environment and returns an LTM individual. This individual can be an *existing* LTM individual (in which case we say that this individual is **recognised**), or a *new* individual. The function thus uses a mixture of immediate perceptual information about the object's location and category, 'contextual' information about the current environment, and LTM information about the location and perceptual properties of known individuals.

At the end of object individuation, a new **WM individual** is created (or an existing WM individual's activation is refreshed). The WM individual is associated (or reassociated) with a FINST, which tracks the individual. The FINST is the 'result' of the whole individuation process.

17.4.3 Cognitive processes involved in reaching-to-grasp

Selecting a subject

The **candidate next subjects function** takes the current LTM environment and the current WM situation and the current allocentric observer state, and computes a pattern of activity over cognitive map locations, identifying interesting things to attend to as 'subjects' for the next episode.[6] I will call this representation the **candidate subjects** representation. This function combines bottom-up input about salient objects/events in the agent's environment, and top-down information about good objects to attend to in the current situation. (This information originates in the WM situation, and is converted into biases towards certain points in the cognitive map by the LTM representation of the current location of objects in the current environment.) I also assume a WTA function which operates on the potential next subjects representation and delivers a 'winning' next subject in a separate layer, which I will call the **current subject** layer. An active unit in this layer is termed the **current subject**. If there is no active unit, there is no current subject.

In the case of a reach action, the subject is an agent, whether it is the observer or a third party. But for continuity with other descriptions I will continue to use the term 'subject' here.

The **orienting function** runs in parallel with the candidate subjects function. It delivers a bias on the saliency map towards the visual features of objects which it would be useful to establish as subjects. It also communicates with the map of candidate subjects, so generate preferences for the corresponding retinal locations. There is a WTA function operating on the saliency map too. By the time a current subject has been established,

---

[6] It probably also takes the totality of motor representations evoked by reflex in the observer, which provide additional biases towards the agent's own location in the cognitive map.

a winning region in the saliency map has been established; i.e. the observer is visually attending to the subject. (If the subject is himself, I will refer to this state as **attention-to-self.**) The establishment of visual attention triggers the **object individuation function** described above, which returns a FINST associated with the subject, and initiates a process which tracks the current subject throughout the remainder of the reach action. The current WM situation is updated with a reference to this FINST.

The **mode-setting function** takes the current subject representation and the allocentric observer location, and returns one of the two action representation modes: **action execution mode** (if the current subject is the observer) and **action recognition mode** (if it is not).

I also assume an **allocentric subject state recognition function**, which returns the allocentric location and orientation of the subject. In action execution mode, the function copies the output of the allocentric observer state recognition function. In action recognition mode, the function copies the output of the allocentric external object state recognition function.

Once a subject has been selected, the bias delivered by the WM situation is altered. Only those WM plans whose first action is consistent with the selected subject remain in contention. The bias these WM plans now imposes relates to the second action in their planned action sequence, since their first action has been achieved.

In summary: by the end of this stage, we have visual (or internal) attention on the subject, we have categorised the subject and activated an appropriate LTM individual, we have established an action representation mode, and we have established the allocentric location and orientation of the subject.

Selecting a target

I will describe the functions in this section on the assumption that action execution mode has been established. Specifying them fully would require each to be specified for action recognition as well. The representations they deliver are the same in each case.

The subject is associated with a set of **motor systems**—units of motor control such as the left hand/arm, the right hand/arm, the system comprising the two hands/arms, and so on. Each of these systems has one (or more) **effectors**: a motor system can be thought of as a device which brings these body parts into stable contact with a target object. Each system is associated with a **motor map** of locations which its effectors can be directed towards, and with a representation of the **current motor state** of its effector, derived through proprioception and an internal forward model.

The **candidate reach targets function** has two components, generating bottom-up and top-down preferences respectively.

The **bottom-up reach targets function** takes as input the **agent-centred saliency map** (which is computed from the retinal saliency map by taking into account the subject's **eye and head position**) and returns a pattern of activity in each of the subject's motor maps. A motor map for a given motor system represents the objects in the subject's perispace, with a bias towards those which are close to its effector, and of the right size.

(Thus a large peripheral object might be strongly represented in the motor map associated with a bimanual motor system, but only weakly represented in the motor map for the left and right hand/arm motor systems individually.)

The **top-down candidate reach targets function** is a visual search function, which biases the saliency map towards the locations of objects which it would be beneficial for the subject to reach for. (These objects are supplied by the revised WM situation, which is now generating preferences for targets). It does this by two mechanisms. Firstly, it maintains a set of **search target categories**, which can be matched against bottom-up results from the object categorisation system. Secondly, it imposes a bias across the saliency map towards the low-level visual features of desired targets, and towards points in the map which object location memory associates with desired targets. These biases percolate down into the motor maps. Likewise, activity in the motor maps feeds back into the saliency map.

I assume there is a WTA function which selects the most active region across all the motor maps. When a winning region is found, focal visual attention will also be allocated to the associated object, allowing it to be categorised. I also assume an IOR function, which allows serial visual search through the candidate target objects until one is found which matches a top-down search target.

The second stage ends when a matching target is found. At this point, we have selected a **current reach target** in one of the motor maps, and we have also established visual attention on this target. In addition, we have selected the **current motor system**, which is the motor system from whose map the winning target was selected. Two things now happen. First, the object individuation function runs again, which returns a FINST associated with the target individual. The WM situation is updated with a reference to this FINST; any candidate WM situations which are incompatible will now cease to be active, and any remaining biases from the WM situation will now relate to the third action in the planned sequence. Second, we initiate a reach to the target object. The **reach motor controller function** takes input from several sources. The selected motor map is associated with a **current motor state** and the current reach target specifies a **goal motor state**. In addition, I assume a specialised visual **effector servoing function**, which provides direct information about the vector from the effector to the target. The output of the controller function is a motor impulse moving the effector towards the target.

Selecting an action category

We have selected a motor system and a reach target, and a reach action is under way, but we have not yet selected an *action category*. As described in Section 2.6.1, there are several actions which we can perform with any given effector on a given target; open-class verbs like *grab*, *hit* and *squash* are distinguished at the motor level by characteristic hand trajectories and by characteristic configurations of the hand in relation to the target. The selection of a motor system and a target object create an attentional environment in which these details can be selected. Once an action category is selected, a second motor controller begins to operate, which can be thought of as *refining* the reach action which is already under way.[7]

I assume there is a set of **action categories** associated with each motor system. Each action category comprises (a) an **effector trajectory**—a sequence of biases on the actual position of the target, (given in motor coordinates, of course), which generates an approach with characteristic dynamics—and (b) an **effector configuration**, which is given in terms of the perceived affordances of the target. The WM situation now delivers a set of top-down biases towards specific action categories, while perceptual processes deliver bottom-up biases.

The perceptual processes assume a selected motor system, together with focal visual attention on the target object. Again, they involve a process of reconstruing objects as environments at a smaller spatial scale. The target object is established as a shape: specifically, as a set of **effector affordances**. There are two types of effector affordance. **Surfaces** represent possible contact points; **opposition spaces** represent possible stable grasps. Likewise, the *effector* of the selected motor system, until now construed as an object moving through space, is now construed as a motor system in its own right. The effector has its own set of motor systems, which I will term **effector subsystems**. There are two types of effector subsystem. **Contact systems** apply unilateral force to a target surface: they are associated with an **orientation**, a **maximum force** (the maximum force which the system can apply without damage) and an **optimal area** to which they should be applied. **Opposition systems** apply opposing forces to a pair of surfaces: they are associated with an **orientation**, a **maximum aperture** (the maximum distance between which opposing forces can be applied) and a **maximum power** (the maximum force can be applied). Each effector system has its own **effector system state** and **effector-centred motor map**, just as at the larger spatial scale. Each effector affordance in the target object is represented in the motor map of each compatible effector system. In each case there is a bias to those affordances which are compatible with the current state of the associated effector system. Top-down biases towards certain effector preshapes also contribute to the selection of an effector system. (For instance, a top-down desire to *slap* will result in a bias towards the contact effector system associated with the open palm.) Within this environment, an effector affordance on the target object is selected, and thereby an effector system is selected.

The current state of the selected effector system, together with the selected effector affordance, provide the input to an **effector motor controller**, which operates in parallel with the reach motor controller. There are two outputs of this controller. One is a motor impulse which moves the current state of the selected effector system towards the selected effector affordance. (For an opposition system, this will involve forces which alter the aperture of the system, and also changes to the orientation of the effector to bring the orientation of the opposition system into line with that of the selected opposition space. For a contact system, it will involve forces which bring the orientation of the surface in line with the selected surface on the object.) The other is a signal which delivers an additional sequence of biases to the location of the the target being used by the reach motor controller,

---

[7] That's an important idea! You need to make more of it—perhaps earlier, in Chapter 2.

to cause the effector to approach the target from an angle compatible with the selected opposition system. (These combine with the biases due to the 'effector trajectory', imposed as part of the selected action category.)

I also assume that there is a special perceptual system for providing input to the effector motor controller, which gives direct information about the relation between the selected effector system and the selected effector affordance. I will call this the **effector system servoing function**.

The final trajectory of the effector onto the target will thus be the product of many factors. It depends fundamentally on a representation of the actual location of the target, computed at the previous stage, and on the impulses provided in real time by the reach motor controller. However, it also depends on the action category selected, which delivers perturbations to the location of the target to achieve particular types of contact. Finally, it depends on the effector subsystem selected as the one to achieve contact with the target, and on the particular surface or opposition space selected on the target, because these impose their own biases on the trajectory of the effector when it approaches the target.[8]

The stable grasp

The combined effect of the reach and effector motor controllers is to bring the effector into a particular kind of contact with the target object. I will assume that this is a stable grasp.

17.4.4 Cognitive processes involved in agent locomotion

Selecting a subject

The process of selecting a subject is the same as for the reach-to-grasp action—see the previous section for details. Once a subject is selected, there is visual (or internal) attention on the subject, we have categorised the subject and activated an appropriate LTM individual, we have established an action representation mode, and we have established the allocentric location and orientation of the subject.

Selecting a goal allocentric location

The idea here: there's a motor map for the *locomotion motor system*, which is given in environment-centred coordinates and takes a target and a trajectory onto

The environment-centred motor controller is learned at points when the subject's *actual* location is the same as his *goal* location.

---

[8] These latter biases can be significant. For instance, if I'm going to slap a target object, then deciding on the *surface* I want to slap has a dominant influence on the way I orient my hand, and therefore on the trajectory my flattened hand takes towards this surface. So much so that maybe it's best to just have a single system, delivering a *combined* hand orientation and effector trajectory, and learn different combinations individually.

17.5 Learning sensorimotor functions

How are all the functions in the previous section learned? I suggest that the key axioms which permit the functions to be learned are often related to certain specific *attentional contexts*. In this section I will describe some of the important axioms which allow sensorimotor functions to be learned, and where appropriate, describe the attentional contexts in which these axioms hold.

17.5.1 Learning the allocentric observer state recognition function

The allocentric observer state recognition function is the function which identifies the observer's own location and orientation in the cognitive map. This function takes as input a perceptual representation of the boundaries of the current environment, and returns a point in the cognitive map and a head direction (see Section **??**), which is invariant over movements of the observer through the environment, and over changes of orientation of the observer in relation to the environment. How is this function learned? As outlined in Section 13.2.1.2, an attractive model is given by Franzius *et al.* (2007), which makes use of the constraint that allocentric representations must only change slowly from moment to moment.[9]

This allocentric observer state function can be computed whenever the observer is in experience mode. It does not matter whether the observer is performing an action himself, or establishing an external subject. If we think in terms of the sequence of functions involved in experiencing a reach-to-grasp action or an agent locomotion action, we can say that the conditions needed in order to train the allocentric observer state function obtain at the very start of this sequence, and at all points thereafter.

17.5.2 Learning the allocentric external object state recognition function

The allocentric external object state recognition function is the function which identifies the location of an attended-to object in the cognitive map. Franzius *et al.* consider this type of learning as well. They show that under conditions in which the animal allocates sustained focal attention to a fixed point in the environment *while itself moving through the environment*, a function which is constrained to deliver a slowly changing representation ends up returning an allocentric representation of the location of the attended object.

For our current purposes, the interesting thing to note is that there is a specific *attentional state* in which the constraint of slow representation changes applies, namely one in which there is sustained attention to a single external object. If the observer switches his attention between different objects in his environment, then of course the constraint of slowly changing representations does not apply. Note also that we have to understand

---

[9] More here, perhaps?

the allocentric representation of location as being the location *of the currently attended external object*; in other words it is deictically specified in relation to the current focus of attention.

Now consider again the sequence of functions involved in a reach-to-grasp action or an agent locomotion action. Note that the attentional state needed in order to train the external object allocentric state recognition function is established from the point when a subject is selected for tracking in either case, and is maintained throughout all subsequent stages. Note that switching attention to the target of a reach action does not break the required correspondences, because the locations of targets are represented in a separate spatial coordinate system, centred on the subject. There is no discontinuity in the location of the subject. Likewise, switching attention to the goal of a locomotion action creates a representation in the 'goal location' layer, not the 'current location' layer.

Given that the function which delivers the observer's own allocentric state can be trained from the very beginning of this sequence, we can generalise, and say that the **allocentric subject state representation function** can be learned at any point after the subject has been selected.

[I should refer to Liddle (2010) at some point in this section.]

17.5.3 Learning the object categorisation function

One important criterion for the object categorisation function is to be able to recognise the category of an object regardless of its current orientation, distance or position in the visual field. Wiskott and Sejnowski (2002) have shown that the constraint of slow representation changes can also be used to establish all these forms of invariance. However, here again this constraint only applies over a time during which there is sustained attention on a single object; i.e. in which a single object is tracked. Here, switches in attention from an agent to a target are important, because the same object categorisation system is deployed first to one then the other. Moreover, recall that the category established by the classification system depends on the spatial frequency of the primitive visual features used as its input; changing this spatial frequency typically result in an abrupt change in category. So the object categorisation function can be learned while there is sustained visual attention to a single object, at a single spatial frequency.

17.5.4 Learning the functions involved in reaching

After having selected a subject, we must select a target and then select an action category. Selecting a target initiates a relatively coarse-grained 'reach' motor controller, and selecting an action category initiates a finer-grained 'grasp' motor controller which refines the motor signal generated by the reach controller. If both controllers do their job, the subject will end up achieving the desired form of contact with a target object in his perispace. I will assume the desired form of contact is a stable grasp.

Both selecting a target and selecting an action category involve many functions. First, there are visual functions which deliver goal representations in the various motor maps.

These can be trained directly at the point when a stable grasp is obtained: at this point, the function which delivers the motor map for the currently established motor system can be trained to map the currently established visual location onto the current effector location, and the function which delivers the map of motor affordances for the currently established effector system can be trained to map the current object shape representation onto the current effector system state. Next there are the servoing functions which deliver direct representations of the difference between the current motor system or effector system state and the corresponding goal state, and the motor controller functions which map current and goal states onto motor signals. These can also be trained from a stable grasp situation; but here some form of temporal credit assignment must be assumed, to permit the discovery of sequences of control signals from perceptual representations which precede the stable grasp state. There is also a clear order in which functions must be learned. The coarse-grained functions which move the hand to the target must be learned first.[10] The finer-grained functions which generate a particular trajectory and hand preshape assume a coarse-grained controller which brings the hand to the target.

17.5.5 Learning the ??? function

Talk about the point at which the subject's goal allocentric location is the same as his current allocentric location. That's the point at the end of a locomotion action. Refer to Hok *et al.* (2007), who have found evidence for a special mode of firing of hippocampal cells, which reflects recognition of having arrived at the intended goal location.

17.5.6 Summary

We can model the cognitive routines involved in locomotion and in reaching-to-grasp as sequences of functions. At each stage, a set of functions establish a certain attentional state, and set up a pattern of connectivity which is maintained until the end of the routine. There are specific points within these sequences where axiomatic conditions hold which allow various functions to be learned. Conditions which allow the allocentric observer state function to be learned obtain from the very start of the sequence. Conditions which allow the allocentric subject state function to be learned obtain once a subject is selected, and persist until the end of the sequence. Conditions which allow the visuomotor transformations, servoing functions and motor controllers for reaching-to-grasp obtain during a stable grasp state. Conditions which allow the object classifier to be learned obtain transitorily, while any single object is attended to at a fixed spatial scale. Conditions which allow the ?? function to learn to navigate through the current environment obtain when the subject's goal location is the same as his current location. (...)

---

[10] This is Tim's MSc project.

17.6 Updates in object location memory

The basic idea here: when you replay a SM sequence to LTM, you create a new temporal individual, and make it the new temporal context, and switch to a new *spatial* context, and then *reactivate* some of the participants in the event, so that they effectively get indexed to a different spatial context (indicating that they have moved).

17.7 Links between sensorimotor and memory representations for individuals

Sensorimotor and memory representations must be connected to one another. In this section, I will summarise how the link is made for individuals, as outlined in Chapter **??**. I will consider the case for events in more detail in Section 17.8.

17.7.1 Attentional actions in memory contexts

When we attend to an object during sensory experience, we perform a sequence of attentional operations, resulting in the evocation of a category complex in IT. The attentional sequence which established the object will be stored as a WM individual, to support later re-attention to the object. In addition, the category complex (together with the current spatiotemporal context representation) may trigger the activation of a LTM individual (in which case we say the object is recognised).

Recall that the IT category complex is a rich structure. In Section **??** I proposed that during sensory experience it can be attended to in a variety of ways, to make different elements of the complex explicit. Attention is implemented by the winner-take-all function in the object category modality, which picks the 'dominant category'. The main operation for exploring the category complex is that BE operation described above, which inhibits the current dominant category, and allows the properties which distinguish the object from the prototypical instance of its dominant category to be evoked one by one.

Recall that when an object is remembered, its category complex is activated. (Remembering an object involves activating a spatiotemporal context different from the current one, and activating a particular LTM individual associated with this context. Activating this pair of representations will evoke a category complex.) I propose that this evoked category complex can be attended to in a variety of ways, just like a category complex which is evoked directly by sensorimotor experience. Thus we can 'inspect' the properties of a remembered individual using the same attentional devices we use to inspect an object present before the senses. Naturally a remembered individual does not generate as rich a sensory representation as an actual object. But I propose that the representation is rich enough to support *actions of reattention.*

17.7.2 The creation of WM individuals during retrieval

Recall that actions of reattention during sensory experience involve WM individuals. The first time an object is attended to, a WM individual is created, to facilitate the process of reattention to this object. How does reattention work for a recalled object? I propose that when an LTM individual is re-activated in a memory context, *a WM individual is created and associated with this individual*, just as happens during sensory experience. Specifically, I propose that when an object is observed in a particular context, we not only learn associations between the context, the LTM individual and its sensory representation, but also between the LTM-individual/context pair and the *WM individual* associated with the object—i.e. the attentional plan through which the object was established. When the LTM individual is retrieved, we also retrieve this associated attentional plan *into WM-for-individuals.*

This proposal solves the problem just noted above: if we want to allow actions of reattention to remembered individuals, we must provide working memory representations of individuals which support reattention. To do this, when we retrieve an individual from memory, we need to create a working memory representation with an appropriate sequential structure, to support the standard repertoire of reattending actions, so that the retrieved sensory complex can be further investigated.

[I need to work out how this fits in with the idea of LTM group environments, which allows LTM to remember the distinction between singular and plural entities.]

17.8 Situation representations

[I think what I'm calling 'working memory' representations of situations in this section I'm now thinking of as 'LTM' situation representations. The current WM situation is really a state of the LTM system, I now think. I now have a slightly different picture of the relationship between WM and LTM, where the pattern of activity over of LTM 'candidate' entities is one handle on the current situation, and the 'currently active' LTM situation is another handle.]

To summarise the previous section: LTM individuals are associated with complex sensory representations, and with working memory representations. Recall from Section 17.3.3.2 that long-term episodic memory also stores representations of each individual 'temporal context' experienced by the agent. In this section we turn to the question of what sensorimotor/WM representations are associated with individual temporal contexts. In Section 17.8.1, I will introduce the notion of a **WM situation** (which can also be called a **WM context**), and define a **situation update function**, which delivers new contexts. (The situation representation, and its associated function, are the most derived elements of the whole model. It's all downhill after this.) In Section 17.8.2 I will introduce the idea that individual temporal contexts in LTM are associated with individual situations.

17.8.1 WM situations and situation types

In the model I present here, a **situation** is an agent's working memory representation of the current moment. It can be thought of in two ways.[11] Firstly, it can be thought of as the representation which supports the agent's decision about which action to execute next. An agent learns what to do next by trial and error: there are various possible actions at any given moment, and an agent must learn by experience which actions lead to positive outcomes. The function which determines which action to perform next is ultimately learned through reinforcement—the model of situations which I will present will provide a framework for reinforcement learning. I am thinking about actions very broadly, to include both substantive motor actions performed by the agent, and attentional actions executed by the agent to locate objects in the world or to observe actions executed by other agents. Thus, for instance, in a particular situation it might be beneficial for an agent to watch another agent, to see if this other agent executes a particular action, or simply to find out what this other agent does.

The second way of thinking of a situation is as a representation of the states which currently obtain in the agent's environment. In the model I present, these states are specified indirectly, in terms of the actions which they afford—in other words, the actions they make possible. There are two classes of possible action. One class concerns possible *attentional actions*. For instance, the state in which there is a cup on the table will be modelled as a possible action of (re)attention to the cup. A possible action of reattention is a WM individual, as will be recalled from Section **??**. A second class of possible actions which define the current state of affairs concerns the *events* which can occur in the situation. For instance, the state in which the agent is holding a cup can be represented indirectly by specifying a set of possible actions which the agent can perform on the cup (namely those actions which require that he is holding the cup); the state in which an observed external agent has just performed an action can be modelled as a set of possible actions which this observed agent might now perform, and which it may be beneficial for the (observing) agent to watch.

The atomic 'actions' which are referred to in a situation construct are not individual sensorimotor operations, but rather *sensorimotor sequence plans*—i.e. WM episodes. Recall from Section 3.3 that plans can compete among one another, can inhibit themselves, and can activate other related plans. In the network I will present, the individual units represent planned sensorimotor sequences. For concreteness, we can assume that each action unit links to a gradient of lower-level action representations in the 'planning layer' of a competitive queueing network which can execute the associated sequence. (For the agent's own motor actions, the first SM operation will be 'attend-to-self'. For actions executed by an external agent A, the first SM operation will be 'attend-to-A'.)

I will model a situation as four banks of action units, as shown in Figure 17.3. In this diagram, each action is represented four times, once in each bank of units, in a way which is reminiscent of the competitive queueing model.

The **chosen action** is the currently dominant PFC sequence plan. Actions at this

---

[11] I also need to relate the WM situation rep to Miller and Cohen's model of the PFC.

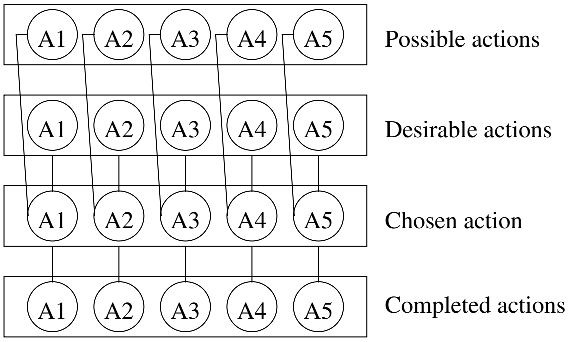


Figure 17.3: The structure of a 'situation' representation

level compete with one another, so that only one can be strongly represented. The chosen plan determines the agent's current behaviour. The chosen action is selected from a bank of **possible actions**, and a parallel bank of **desirable actions**. (The function which combines desirability and possibility will be considered in more detail in Section **??**. For the moment, we can think of it as a function which computes the 'expected benefits' of each action in the agent's repertoire, so that the chosen action is the one with the greatest expected benefit, perhaps chosen stochastically to enable a component of exploration.) I will refer to the possible actions and desirable actions layers as the **planning layers** in what follows.

17.8.1.1 Operations triggered by completion of the current plan

When the currently dominant plan is completed, it must be removed from the chosen action layer, so that another plan can be established there. I assume a mechanism for recognising that the current plan has been completed, which triggers four distinct operations. One of these is the operation of inhibiting the action at the 'desirable actions' level, which is similar to the self-inhibit circuit that operates in the competitive queueing model at the next level down. This inhibition-of-return operation prevents the same action being executed again.[12] Another operation triggered by a successful plan is an operation of *internally rehearsing* the sensorimotor sequence associated with the completed action, for the purposes of storing it in episodic memory—as was discussed in Section 3.8.1.3. (This operation must obviously occur *before* the inhibition of the associated desirable action, so that the just-completed plan is rehearsed before it is removed from the dominant-action layer.) A third operation triggered by a successful plan is the operation of **updating the current situation representation**, to take into account the consequences of the just-completed action. The situation update function will be described in detail in Section 17.8.1.2—its basic effect is to update the 'possibility' and 'desirability' of each action in the agent's repertoire, to reflect the changes in the world brought about by the newly completed event.

---

[12] If this operation is not performed, the agent will repeatedly execute the same action. Iterated actions can thus be modelled by selectively disabling this operation—see Section **??**.

The fourth operation is involved in *learning* an improved version of the situation update function, to be used next time the agent is in a situation similar to the one he has just encountered. The fact that the action has just been successfully accomplished in the just-encountered situation can be used as a training instance for the function which decides which actions are possible in which situations. It can also be used to train the function which specifies the desirability of the current action in the current situation. I will describe both of these operations in more detail in Section **??**.)

17.8.1.2 Completed actions and situation updates

Active units in the **completed actions** level denote actions whose effects still have behavioural relevance for the agent. For instance, if the agent is in the middle of making tea, and completes the action of boiling the kettle, the sets of possible and desirable actions change in certain respects, as a function of the newly completed event. Thus completed actions bring about *updates* to the sets of possible and desirable actions.[13]

The updates which are brought about are often quite complex. I will model them by introducing a function, the **situation update function**, which takes the set of completed actions onto a set of *updates* to the sets of possible and desirable actions.[14] Thus, a new completed action can make other actions in the agent's repertoire more or less possible, and more or less desirable. For instance, if the agent has completed the actions of boiling the kettle and putting tea into the pot, the two completed action representations might have the effect of increasing the desirability of the action of pouring the boiled water into the pot. (Note that it does not matter which order the two prerequisite actions are carried out in.)

The behavioural relevance of a completed action can change over time, or as a function of subsequent events. For instance, when the agent has finished making his tea, and is drinking it, the completed kettle-boiling and teabag-adding actions no longer drive his behaviour. We must therefore assume another function, which deactivates completed action representations, and which undoes the effects they had on the planning layers. The deactivation mechanism might involve decay over time (with different time periods for different kinds of action), or something more complex, with sensitivity to the goals which actions serve.[15] The important thing is that the changes which are made in the planning layers when the completed action becomes active are undone in measure of its deactivation, so

---

[13] I should definitely mention Brass and Haggard (2010) here. They have an idea that the experiencer 'evaluates' the consequences of any actions she performs, and that this evaluation (positive or negative) makes them 'more or less available in *future* situations.' They don't have a formal model of situations, but the idea that a situation includes a set of evaluated potential actions is certainly implicit in their account. (In their model, the evaluation is done in anterior insula.)

[14] Actually I think it's better if the situation update function takes the chosen action, *at the point when it self-inhibits*. If it takes the completed actions, it's not clear how to update if the completed actions layer changes only partially (which is typically how it changes).

[15] I have in mind something like the mechanism which knows when to expect a reward or punishment; c.f. O'Reilly *et al.* Beyond the point of an expected reward, perhaps we can deactivate the states put in place to bring about the reward.

when it is no longer active, it no longer has any impact on these layers.

[I think some notion of *nested situations* might be sufficient (and more general) as a way of dealing with completed actions. When you're making tea, and you have boiled the kettle you're in a new situation, but it's still a situation *within* the wider 'making tea' situation. After you've gone through several nested situations (which are like nested environments), and you achieve the final goal of making tea, you move out of the *wide* situation (making tea) as well as out of more narrow situations. That sounds more general.]

17.8.1.3 Episodes(?) and situation types

One way of characterising a situation representation is by looking at the patterns of activity in the possible and desirable actions layers (i.e. in the planning layers). Can these patterns of activity be decomposed into separable components? At a rough approximation, we can attribute the activity in these two layers to different independent sources. Some activity relates to the currently active *completed* actions. Other activity relates to plans which are currently *ongoing*, which provide an independent component of activation in the desirable actions layer. Within each of these sources, there is additional structure. There may be several completed actions which have quite independent effects on the planning layers. Likewise, the agent may have several quite independent ongoing plans, which contribute relatively independently to activity in the desirable actions layers.

It is useful to think about the planning layers as a representation which **categorises** the current situation. Just as IT can hold several different categories for an object, and represent these categories simultaneously, so a situation can represent several simultaneous patterns of activity in the planning layers. Just as IT has a winner-take-all function which identifies the dominant object category at any given time, so a situation has a function which selects a single action as the one which drives the agent's immediate behaviour. Just as IT has a function for inhibiting the currently dominant category, so a situation has a function for inhibiting the currently dominant action plan—which by default operates when the action plan has succeeded, as described above in Section 17.8.1.1. And just as the self-inhibit function in IT generates a new category complex, so the event of a plan succeeding brings about a new pattern of activity in the planning layers in a situation.

Given this analogy, note that it is possible to think of individual plans in a situation representation as associated with **situation types**. When a plan is under way, it is associated with a certain pattern of activity in the desirable actions layer. When a plan succeeds, it results in a reasonably well-defined change in the activity of the possible and desirable actions layers. We can thus talk about the *type* of situation in which a certain plan has just succeeded, and has behavioural relevance—there are certain patterns in the planning layers which are characteristic of this type of situation. An **individual situation** may be of several different types simultaneously, just as an individual object in the world may be of several different categories.

To summarise: when a planned action succeeds, it updates the current individual situation by representing it as a situation of a certain type, in which a new distinctive pattern of possible and desirable actions is added to the patterns already present in the planning layers.

17.8.2 Individual situations and episodic memory contexts

An agent always has an active situation representation in working memory. This representation is updated after every action he performs, and after every event or state he observes. At the same time as the agent is acting and experiencing the world, he is creating a memory representation of it. I propose that the updates of working memory situation representations are coordinated with updates of **temporal context** representations: each updated situation representation is associated with a new temporal context representation in the episodic memory system, which we can call an **LTM context**, or a **temporal context**, or an **individual context**.

Recall that episodic memory involves two types of context representation: a spatial contex and a temporal context. During experience of the world, an agent has to create new individual temporal context representations, and associate them with the episodes he experiences. Later, the agent can re-activate an older temporal context representation, to re-create, or re-enact, events experienced in the past. I have not been very clear about what a temporal context is yet. In this section I will define it in more detail.

In the model I propose, an individual temporal context, denoting a particular time in the past, is in many respects similar to an LTM individual representation—i.e. to our memory representation of an individual object. When a new individual situation is established in working memory, a brand new temporal context is created, and the two representations are linked by Hebbian associations. Later, during recall, the WM situation can be used as a retrieval cue. If the WM situation is re-evoked as a retrieval cue, this has the effect of reactivating the temporal context in the episodic memory system, and of entering a mode in which events and states associated with this reactivated context can be explored. Conversely, if the temporal context is re-activated, this has the effect of activating the associated WM situation.

Recall that when an LTM individual is reactivated, this activates a *complex* of object categories in IT and associated areas, which can be explored in different ways by different attentional operations. I propose that when an individual temporal context is reactivated, this activates a complete situation, with a complex structure of possible and desirable actions, which can likewise be explored in different ways. I will look at some of the things we can do in the remainder of this section.

17.8.2.1 Linkages between temporal contexts and episodes in episodic memory

We have already seen how LTM individuals are linked to individual temporal contexts. How are *episodes* linked to contexts? The story here can be quite simple, since there is only one episode that occurs in each context: an episode is what creates a new context. I will assume that individual temporal contexts are directly associated with each other in chains, so that when one is active, there is a special kind of operation roughly corresponding to the question ‘what happened in this context?’ which activates the next context.

I will assume that activating a *pair* of adjacent context representations has a special effect, which is to retrieve a representation of the event which caused the transition between them.[16] As was described in Section 3.7, events are represented in episodic memory as sequences of sensorimotor operations. In Section **??** I proposed that event-depicting sensorimotor sequences are linked to spatiotemporal contexts: the idea was that activating an individual context will directly trigger the sequence representing this event. This idea now needs to be a little refined. I now propose that if we activate context $C1$, *and then execute the 'what-happened?' operation* to activate the next context $C2$, this *pair* of context representations will trigger the event-denoting sensorimotor sequence. I also propose that once the sequence is completed, the original context $C1$ is inhibited. If we want, we can now execute the 'what happened?' operation again, to generate the event that occurred in context $C2$, and so on.

Recall from Sections 3.6.4 and 3.7 that the event-denoting sequence which we store in episodic memory makes reference to *LTM individuals.* For instance, the sequence representing *The man grabbed a cup* would first involve activation of the LTM individual associated with the man, and then activation of the LTM individual associated with the cup, and then activation of the premotor *grab* schema. Recall from Section 3.8.2.6 that this sequence is retrieved into (episodic) working memory, which recreates a sequence plan representation similar to the plan representation created in working memory when the event is first experienced. Recall also that when an LTM individual is activated in an episodic memory context, this triggers activation of an associated WM individual. The working memory representation of the event interfaces with the working memory representations of the individuals which the event involves, in ways which I will discuss in Section **??**.

17.8.2.2 Attending to objects in memory contexts

I am not obliged to ask the 'what happened?' question when I evoke a memory context. Another way I can explore the context is to execute one of the possible *attentional actions* which the WM situation associated with the context permits. These actions do not update the memory context; they simply explore it, in the same way that attentional actions can explore the sensory representation evoked in IT by an activated LTM individual.

This proposal can be thought of as a proposal about the differences between the way events and states are encoded in episodic memory. Events are encoded as sequences which link two memory contexts, as just summarised. I suggest that states are encoded quite differently, in situation representations held in working memory. Recall that an agent has two separate ways of representing 'the current moment' in memory: the current moment is represented in the episodic memory system as an individual *temporal context*, and in the working memory system as an individual *situation.* Recall also that these two memory representations are linked together: the new context is associated with the new individual situation. If this temporal context is subsequently reactivated, it will reactivate the

---

[16] This idea of representing an event as a pair of precondition state and consequent state is similar to the model of 'event nuclei' in Moens and Steedman's model of tense and aspect. I want to make something of that connection eventually.

associated working memory situation. As discussed in Section 17.8.1, the working memory situation represents the states which currently obtain as a set of *possible actions of reattention*—in other words, as a set of *WM individuals*. For instance, the state in which there is a cup on the table is represented by the WM individual which supports reattention to the cup.

To flesh out this notion of stative memory, we must also specify that whenever an agent executes an action of attention, the updated WM situation must be re-associated with the current episodic memory context. Each new action of attention enriches the WM situation; and each of the enrichments must be linked to the episodic memory context, so that when the context is retrieved, the complete situation is also reactivated.

What this means is that while episodic memory is specially configured for sequential retrieval of the structure of events—so that they are recalled in the order in which they happen—stative memory allows access to a wide range of states that obtain at the currently active temporal context. The attentional actions which are executed are under the control of the agent *at the time of recall*: they are not determined by events occuring at the time of storage. This idea is at the heart of the model of questions and answers which I develop in Section **??**.

17.8.2.3 Retrieval of event sequences and the situation update function

Consider again the scenario where we reactivate a past episodic memory context $C1$ and then ask *what happened?*, to activate the next context $C2$, and recall the event E which linked the two contexts. It is interesting to consider this operation from the perspective of working memory situations. Every episodic memory context is linked to a WM sitation: they stand in a one-to-one relationship to each other. Given that we have a situation associated with $C1$, and we have retrieved the event which takes us from $C1$ to $C2$, we actually have a second method for evoking the situation associated with $C2$: we can apply the *situation update function*, as we would do during actual experience. This allows us to generate *expectations* about the updated situation. There are two things which follow from this. Firstly, we don't have to encode everything about the consequent situation in gory detail—expectations can do a lot of the work. Secondly, anything which occurs in the actual situation $C2$ which is not expected is particularly interesting, and warrants special attention. This idea should feature in the model of coherence relations like that signalled by *but*, and also in the model of how deviations from expectation are encoded in memory.

[Actually, this idea of expectations also applies during *experience*. Firstly, I can rely on a mature situation update function to take lots of things for granted about the new situation which I find myself in. Secondly, if experience ever shows me cases where the situation update function *can't* relied on, these cases are valuable for learning an improved version of the function. (And in the meantime, their exceptions should be encoded explicitly in episodic memory.)]

[I should say something about how the amount of new 'situation-specific' learning which is done should be dependent on how predictable the situation update is. If it's fully predicted, no LTM change needs to be made at all. If it's really surprising, it needs to be strongly encoded in LTM as an individual (sparse) situation. This probably relates to the role of neurotransmitters like dopamine as modulators of synaptic plasticity.]

17.8.2.4 *Can* and *should*

The situation representation contains explicit representations of the actions/states which the agent desires, and those which the agent deems possible. Recall from Chapter 5 that we have already proposed a special method by which the agent can *rehearse* the plans associated with a desired action, even if the plan has not been successfully executed. The proposal was that the verb *want* denotes an operation which establishes a special mode, in which the agent's most desired plan is selected and rehearsed. (Before entering this mode, the most desired plan inhibits itself, so that the agent does not read out the plan to read out the desired plan.) The syntactic fact that *want* takes a nonfinite clausal complement was then interpreted as a corollary of this combined self-inhibit-and-plan-read-out operation.

The situation representation introduced in the current sections allows us to define the self-inhibit and plan-read-out operations more clearly. We can say that *want* denotes a special operation which (i) causes the current plan to succeed straight away, triggering it to be read out and self-inhibit; (ii) establishes a mode in which actions compete on the basis of desirability alone, and in which the read-out operation to be triggered as soon as an action is chosen.[17]

The situation representation in fact provides a framework for a number of special operations of this kind. The modal verb *should* can be thought of as expressing something similar to *want*—at least in its 'practical' interpretation, as a description of what is best for the agent.[18] The modal verb *can* can be thought of as denoting an operation of the same kind, but one which restricts competition to the 'possible actions' layer.

An interesting thing about the present analysis of situations in memory is that the agent is just as able to perform the *want*, *should* and *can* operations *in a remembered situation*, as well as in the actual situation. So the agent can 'explore' actions he didn't do, but *could have* or *should have* done. A treatment of counterfactual possibilities like these is essential to any treatment of modality in natural language (see e.g. Steedman's model of tense and aspect).

I want to talk more here about how the banks of possible and desired actions relate nicely to the notions of accessibility in modal logic.

---

[17] It's important that we *read out* the 'I want' plan before it's self-inhibited. The idea is that you're reading out that part of the sentence before you even know how it's going to be completed. Which might have testable consequences.

[18] *Should* also has a 'deontic' interpretation, in which it describes what the agent's 'duty' is. To capture this interpretation, we can envisage another bank of units in the planning layer, which we could term **prescribed actions** which encode the agent's socially instilled responsibilities, and which provide a third ingredient to his decision-making processes. Deontic *should* would then denote an operation similar to *want*, but where competition is restricted to the prescribed actions layer rather than the desired actions layer.

17.9 Nonstandard situation updates: LTM retrieval, conditionals and unexpected events

During normal experience, the situation update function controls the transition from one WM situation to the next. However, there can also be more abrupt discontinuities in the sequence of WM situations. These can occur when the agent disengages from the current WM situation to recall a previous situation or to consider a hypothetical situation. They can also occur when an unexpected event happens. I will consider the former cases in Sections 17.9.1 and 17.9.2, and the latter case in Section 17.9.3.

17.9.1 Retrieving LTM situations

During an agent's experience, a brand new LTM situation is created each time an action is completed and a new WM situation is created. The new LTM situation is linked by Hebbian associations to the new WM situation. In this section, I will discuss how a LTM situation can be retrieved.

'Retrieving a LTM situation' involves bringing its associated WM situation back into working memory. It involves moving into Tulving's 'memory mode', where the contents of working memory reflect a previous time, rather than current sensory experience. What kind of memory cue is able to retrieve a LTM situation, and/or cause the activation in working memory of a situation representation other than the one which actually obtains? In this section I will focus on cues which are presented linguistically.

Consider the following sentence:

(17.1) When you got home, the cup was on the table.

I assume the effect of this sentence is to cause the agent to revisit an earlier situation (having got home), and to update his representation of that situation by adding a new fact, in the same way he would have updated it had he noticed the fact at the time. I propose that the sentence denotes a sequence of SM operations, as usual, but that some of the operations have a special role in constructing and issuing a cue to LTM. The operations fall into two stages: first the accessing of the earlier LTM situation, and then the adding of a new fact.

In the former stage, the first operation is to clear the agent's current WM situation representation, to create a blank slate on which a cue to LTM can be prepared. Next, a new WM plan associated with the action 'I get home' is evoked as the 'chosen action' in the situation medium, and this action is assumed to succeed, triggering the situation update function. We now have an unusual WM situation representation: it just represents the consequent state of a single action. WM situation representations built during actual experience are usually much richer than this, reflecting the many different facets of the situations we find ourselves in. The point about this simple representation is that it can function as a cue for retrieving a LTM situation. When the agent actually did arrive home, he created a new LTM situation, and linked this by Hebbian association with the rich WM situation representation he was currently maintaining. This rich WM situation had many components—but one of these components was the one which has just been recreated in WM as a memory cue. If this is a strong enough component, then it will reactivate the LTM situation. This in turn will reactivate the complete WM situation. In summary, a partial WM situation activates an LTM situation, which in turn reactivates a complete WM situation. The agent is now re-experiencing something like the state he was in when he arrived home.

In the second stage, some new material is associated with this re-experienced context. A standard idea in theories of sentence semantics is that any incoming sentence is interpreted as an update to a representation of 'the current discourse context'. By default, the current context is the context which results from processing the previous utterance in the discourse (see e.g. Kamp and Reyle, 1993). A *when*-clause selects an arbitrary new context as the current context; however, the material presented in the main clause is added to this context in the usual way. In our model, we can assume that Hebbian associations are created between the current LTM context (however it was established) and the WM situation evoked by the main clause. This amounts to proposing that the *when*-sentence in Example 17.1 updates the agent's representation of the re-experienced situation in the same way it would have been updated if he had noticed the cup on the table when he came in.

The above account of *when* is very simple. Linguistic accounts of *when* are a good deal more sophisticated. For one thing, the **tense** and **aspectual type** of the subordinate clause and main clause strongly influence the nature of the relation signalled by *when*—see Moens and Steedman (1987) for an accessible introduction). For another thing, the agent may not have an existing representation of the situation presupposed by the *when* clause. In this case the agent has a choice: he can trust the speaker and **accommodate** the presupposed situation (see Stalnaker, 1973 and much subsequent work), or reject the speaker's utterance altogether. My aim here is just to describe the cognitive operations which disengage the hearer from his current sensory context and revert to an earlier context, or a newly created one.

17.9.2 Conditionals

The subordinating conjunction *if* is in some ways like *when*. A conditional sentence using *if* establishes a new context, unrelated to the current sensory context, and associates this new context with an action to be executed, or a conclusion to be asserted, if this context ever occurs. An example is given below.

(17.2) If John comes in, say hello.

The main difference between *if* and *when* is that a sentence using *when* presupposes that the material in its subordinate clause will occur, or has occurred, while a sentence using if makes no such presupposition. *When John comes in* presupposes that John will come in; *When John came in* presupposes that John came in at some point in the past.

In my model of *when*, the clause introduced by *when* functions as a cue to episodic LTM. This explains its presuppositional force; it accesses a situation which has been encoded in episodic memory. (Note that episodic memory extends to scenarios which are predicted to occur in the future—something which Tulving is at pains to emphasise in his model of episodic memory.)

I propose that a sentence using *if* denotes a sequence similar to that denoted by *when*, but lacking an operation of accessing LTM. The sequence has two stages, as for *when*. In the first stage, the current WM situation is cleared, and the counterfactual situation is established as the current WM situation. But there is no LTM access operation. This means that the WM representation *by itself* functions as the 'current context'. In the second stage, the consequent of the conditional is activated as the dominant WM plan, and Hebbian associations are created between it and the current context, just as for *when*. But since these associations are all within the medium of working memory, the effect is simply to change the agent's strategies, rather than to assert anything new into LTM.[19][20]

To take an example, consider Example 17.2 again. The sequence denoted by this sentence has two stages. In the first stage, the WM situation is cleared, *John comes in* is activated as the 'chosen action', the situation update function is executed to produce a representation of the state which results from this action in the planning layers, and *John comes in* is cleared from the chosen action layer. The planning layers function as the 'current context', and remain active. In the second stage, the action *say hello* is activated in the chosen action layer, and Hebbian connections are made between this action and the active situation representation in the planning layers. The effect of this is to make it more likely that the agent performs the action *Say hello* in the situation which results from having noted that John has come in.[21]

This account of the difference between *when* and *if* trades on the dual role of working memory in the current model. Working memory is used both to form cues to LTM, and to learn general action strategies. A working-memory-based model allows an interesting account of the similarities and differences between *if* and *when*.

17.9.3 Unexpected events

Another interesting use of *when* is in a narrative, to introduce a surprise new development. For instance:

(17.3) I was walking through the park when a man jumped out from behind a bush.

---

[19] You need to emphasise this point. A conditional—i.e. a general rule—denotes a connection *purely within working memory*, without any reference to LTM. It is this which gives conditionals their rule-like, or universal flavour.

[20] You should also note that the idea that general rules are maintained in PFC-based working memory is fairly common currency in computational models of cognitive function—see especially Miller and O'Reilly.

[21] Actually, I don't think that's right. I prefer if in the first stage, we activate *John comes in* in the 'chosen action' layer, then in the second stage we activate *Say hello* in the *desirable actions* layer, and learn a link between these two which will trigger a similar update after subsequent *John comes in* actions are completed.

It is hard to explain how this use of *when* relates to the operation of accessing LTM. My proposal is that the link concerns the way the WM situation is updated. When I am walking through the park, I have a representation of the current situation which supports a particular chosen action. I assume that this action does not take up all my cognitive resources, and that I can continue to monitor the world while I am performing the action. During ordinary walking, I might observe several things - for instance, the presence of other people on the footpath, the colour of their clothes, and so on. Each observation briefly features in the 'chosen action' layer, and a situation update is performed. In most cases, the situation update function is the identity function: the observation self-inhibits without any changes to the planning layers, because it has no consequences for our behaviour. But when the observation is important—for instance, a threatening man appearing—the situation does change.[22] The consequent state of *a man jumps out from behind a tree* dominates the existing situation in the planning layers, and the agent's subsequent actions will be determined primarily by this state. (He might decide to avoid the man, or confront him, and so on.) Note that the action also remains active in the 'completed actions' layer. This is important if later the situation returns to normal: in this case, the action is inhibited in the completed actions layer, which has the effect of inhibiting those components of the planning layer which relate to this action. The dominant situation remaining in the planning layer is the one associated with walking through the park, so the agent will continue with this action.

17.10 Hierarchical structures and hierarchical operations in situation transitions

These relate to different discourse structures, I think.

17.11 Reinforcement learning in the WM architecture

17.11.1 Temporal difference learning

Refer back somewhere else?

17.11.2 The actor-critic framework

17.11.3 An implementation of an actor-critic algorithm in WM

---

[22] This situation is very similar to Braver and Cohen's scenario of the agent's current PFC stage being updated. It's a situation where an unexpected reward (or punishment) is predicted, which prompts an update to PFC.

Chapter 18

A new model of LTM, and its connection to the sensorimotor system

This is a summary of some new ideas.

18.1 A framework

- There is a system of **candidate units**, a system of **current LTM units**, and a system of **current sensorimotor units**, which stand in a 1:1 relation to one another.[1]
- The current LTM units are hippocampal/temporal; the candidate units are prefrontal. The current SM units are—? Some combination of parietal and premotor??
- There are different banks of units: **individuals**, **environments**, **locations**, and (perhaps) **situations**.[2] These tell us about ways of construing the world, rather than directly about the world. (Because what we intuitively think of as an 'object' can be experienced as an individual, as an environment or even as a situation.[3]
- There are certain set relationships between these things.
  - An environment has associations with candidate individuals, and with candidate locations.

---

[1] The implication here: an *update* to the 'current SM unit' needn't occur at the same time as an update to the 'current LTM unit'. I'll play with the idea that they happen at different times. (Maybe *when* and *if* tell us about updates within only one system?)

[2] What about property assemblies? An LTM individual links 1:1 with a property complex, which is a set of candidate property assemblies.

[3] Actually I'm not quite clear about that last one.

  - An individual has a single ‘special’ association with one particular environment, which represents it ‘when it is established as an environment’. (In fact, I think there are some individuals which have associations with *several* environments—as in the case of an agent, which when established as an individual can allow transitions to several different alternative (motor) environments.[4])
  - When you activate an environmnent, you also *enable* a set of bidirectional associations between individuals and locations. These are ‘object location memory’. (At least, in one of their forms they are.)
  - In the network, there will be a useful notion of **overlaid distributed representations**. This is a classical neural network idea, but I want to make reference to it. The difference should be that there are localist units (the ‘current units’) which link to *specific* distributed representations, and allow a single representation from this overlaid set to be (temporarily) evoked by itself. (E.g. the ‘property complex’ layer can exist in two modes: either it’s representing a single individual, or it’s representing a set of possible individuals, and being used to help select one of these—after which it transits to the first mode.) (E.g. There may be several notions of ‘the current situation’, some which come from a representation of the current physical environment we’re in, and the objects which are known to be in it, some which come from a representation of the semantic associations of the environment, some which come from our current goals; these all deliver independent biases on the set of salient environment-centred locations, which are overlaid on top of one another.)
  - This notion of overlaid distributed representations should also be the way we represent whole environments, or situations: they let us compute a probability distribution over the next action, or the next 2 actions, or the next 3 actions, which is sufficient to suggest a single sequence of actions. At the same time, if this sequence of actions is unsuccessful, or doesn’t bear up to sequential scrutiny (look-ahead), we can inhibit it, and the environment/situation will deliver us the next-best one.

- There is a notion of well-defined updates within the system. There can be different sorts of update.
  - Within the current environment, you can select the most salient individual as the ‘current individual’.
  - Within the current environment, you can use IOR to switch from the current ‘current individual’ to another one.

---

[4] There’s something here about locations within these alternative environments all competing on an equal footing, so when you select a location, this automatically selects the environment it’s in as well. (Or perhaps better: you select a *pair* of location and environment which happen to be ‘compatible’, i.e. the former is ‘in’ the latter.

  - The current individual can be reconstrued as an environment, and become the 'current environment', within which there are various locations.
  - We can 'pull back' from the current environment, to the environment which contains it as an individual.

- Updates can happen in the SM system as well. (Actually, I suggest these updates *always happen first*; see below.) An update in the SM system results in some action and/or perception actually being done. These updates all involve the selection of a new 'current SM unit'—either a SM individual, or a SM environment (or both).
  - These updates don't necessarily synch up directly with LTM updates. At least, we don't always establish a new LTM *situation*. We always update the 'current subject place' and 'current subject orientation' (environment-centred reps), and we sometimes update 'object location memory' (if we happen to establish an object). But if we attend somewhere and there's nothing there, I suggest we don't update object location memory; and if we try an action and nothing comes of it, I suggest we don't update the LTM situation.
  - I suggest that when we select a new WM situation, that basically starts a motor action. I suggest that when that motor action reaches its consequent state, this acts to trigger an update of the current *LTM* situation. Thus 'an update of the temporal index' (which is what 'events' are supposed to produce) is in fact a *pair* of updates, which are *staggered*, and occur one after the other.[5] In the first phase, we have a current WM situation which 'matches' the current LTM situation. In this context, there's a pressure on the WM situation to update. (Because the LTM situation generates activity in the layers of candidate units, which pushes us towards a different current WM situation.) So eventually, we get a *new* current WM situation. Now we're in the second phase: the current LTM situation *doesn't* match the WM situation. I suggest this means 'an action is taking place'. At the end of the second phase, we attain the consequent state of the action. This is a *sensory* state. *The end of the action is also a piece of sensing of the world.* The world is now different (because the action changed it). So we have to update to a new LTM situation. We represent the new situation axiomatically as 'the situation brought about by the action', as in the situation calculus. This makes the new current LTM situation once again 'match' the current WM situation—and we're back to phase 1.

- I suggest that updates are to **stable states** of the system. But we typically need to perform *several* operations to get from one stable state to another one. I'll call these operations **intermediate operations**. For instance, completing a motor action is a 'stable state' in the 'situation' domain. But doing this requires intermediate

---

[5] This is reminiscent of Moens and Steedman's 1988 idea that an 'event nucleus' has a preparatory state as well as a consequent state—i.e. an event contains something that matches the start and something that matches the end state of the episode it describes.

operations—preparatory attentional actions, to set up the appropriate coordinate systems. We can backtrack on intermediate operations if they don't give the right results. Another example: establishing an object is a stable state, so that's an update point (where the LTM system can be brought into synch with the SM system). But establishing an object involves a few intermediate operations; attending to a location (and finding a good salient point), categorising what you see (and getting an informative property complex).

- There are some systems of sensorimotor units that connect to the world in an axiomatic way. Units which are environments each have a **sensory** component, a **transformation** component and a **motor** component. Units which are individuals I'm not sure about.
  - **SM individuals**. A SM individual is a combination of a salient location and a 'clear' property complex. (The idea is that not every location is good enough: it has to be salient. And not any activity in the property complex is good enough: it has to be strong enough to exceed some threshold. I.e. the individual has to have some reasonably distinct visual property/properties.) The property complex is just some rich distributed representation of visual properties. A SM individual is a SM state which is sufficient to activate an LTM individual (through 'recognition' or the creation of a new LTM individual). Pt: there need to be several components to this state: minimally, location and property complex, but also perhaps something to do with cardinality (to indicate whether the property complex should be related to the individual directly, or to its associated environment, if it's a 'group'.)[6]
  - **Top-level SM environments**. The sensory component of a top-level SM environment is an unfiltered sensory manifold: to be concrete, I'll say it's the output of primary visual cortex across the whole visual field. (Obviously there will be other modalities as well.)[7] The transformation component is a function which maps this manifold (instantaneously) to a representation of the location of a 'subject' (either the observer himself or an external object attended to as a subject) in the observer's current environment. The motor component comprises the system controlling the observer's direction of gaze, and the system controlling locomotion.[8]

---

[6] SM individuals relate to what I called 'WM individuals' in Chapter 10. WM individuals at their most basic are associations between property complexes and locations. I was suggesting that they are 'attentional sequence plans', so that when you attend to a location you know what the rest of the attentional sequence should be; and when you activate an LTM individual top-down, you know what attentional sequence should help perceptually establish it. I'm not sure how to read the new notion of a 'SM individual' as a planned sequence. The main thing is that it should be something which can be *replayed* (in simulation). One idea: the *sequential* component to the replayed SM operation is actually implicit in the order in which units must become selected in the *LTM* system.

[7] Perhaps the unfiltered sensory manifold can be the saliency map. I.e. it's not *completely* unfiltered; but it is computed from the whole visual field, including the periphery.

[8] The observer can access information about the locations of other objects in the environment, because

- An **agent-centred environment**. Assume we have an LTM individual which happens to be an agent. If we establish this individual as an environment, what we get is a special **agent-centred environment**. The sensory component of this is a **filter**, which is applied to the currently active sensory manifold, to deliver a **filtered sensory manifold**.[9] (The filter represents something like 'the agent's peripheral vision': it's not *focal* attention on one thing; rather, it 'weights' the candidate entities in the agent's environment.)[10] Spatially, the filter gives extra weight to salient locations which are at the fovea, within the field of view, close to the body. The transformation component of an agent-centred environment uses info about the agent's eye, head and body position to map from environment-centred coordinates to agent-centred (eye/head/body-centred) ones. The motor component of an agent-centred environment is again gaze direction and locomotion. (Note: while the top-level SM environment makes use of gaze direction and locomotion, the point is that it learns to *abstract over* these actions; while the agent-centred environment *uses* gaze direction (& probably locomotion too) to orient the agent towards bits of the environment which are especially interesting.)
- **Focal SM environments**. Each 'location within' an agent is a focal SM environment. The sensory component of a focal SM environment has two parts. Firstly there's the agent-centred environment. Secondly, there's a haptic system, which delivers information directly to a specific motor system, in a coordinate system specific to a particular focal SM environment (i.e. specific to a particular motor system). This coordinate system is defined by the relationship between the particular motor system in question and the location and coordinate system of a *selected object*. (Now I'm talking about a *single*, *actually selected* object.) The 'transformation' component of a focal SM environment maps the agent-centred filtered sensory manifold into the coordinate system of that focal SM environment, so that it can use distal as well as proximal/haptic information

---

this is stored in *object location memory*: some mixture of LTM and WM. This form of memory indexes several objects to the environment as 'candidate individuals', and each of these activates a 'candidate (environment-centred) location', so we have a gradient of candidate locations we might choose to orient to. When we *select* a winning location, we do so first in the SM system, resulting in an action of attention to this location. (Which only 'succeeds' if we find an actual salient bottom-up location.) Once a salient location is found, this reinforces ('locks in') the associated *candidate* location, which changes the patterns of activation in the other 'candidate' layers. In particular, it gives a much reduced shortlist of candidate LTM individuals; namely those we expect to find at the selected location. When we then get a property complex in IT, this further reduces the set of candidate LTM units—perhaps to a single unit—in which case we have *recognition*. (It's also possible that we need to establish the salient region as an environment before we do classification. In this case, the candidate LTM individuals will activate their associated environments, and *then* we get a template which allows recognition. Note that the SM operations are paralleled by refinements in the active sets of candidate units right down the line.)

[9] Or perhaps a filtered *saliency map*.

[10] If the agent is the observer himself, this is a 'basic' representation of the world, that's computed directly. But perhaps it's not *evolutionarily* basic.

  about the object. I think it also somehow maps the environment-centred LTM information about the locations of LTM individuals into The 'motor' component of a focal SM environment again has to do with an agent's gaze system, but also with a specific motor system of the body. The gaze system is only indirectly relevant: we don't have to be gazing at the selected object, but we probably will home in on it with a saccade and/or a head movement and/or a body movement.

- An agent is always *in* an environment. When it comes to actions, he only has two choices: (i) he can act 'directly' on the environment (by locomoting through it, or sampling it perceptually; (ii) he can act on an *object* in the environment, which involves first establishing the object as a focal SM environment, and then in this new context acting 'directly'. (Note that the latter option in some sense includes the former option as a component, because you are sensing the environment when you act on an object within it.)

- 'Infinite recursion' is possible within the LTM system.[11] In the SM system, of course, there are a limited number of 'basic' hierarchical levels, but there's scope for changing the way the SM system is deployed to the world, allowing for something like recursion here too.

  - In the visual system, the 'basic' hierarchical levels are to do with the different spatial scales vision makes available. The broadest scale encompasses the whole visual field, and delivers the position of the subject in the observer's local environment. The next scale down offers locations in this environment, which are either individuals or sub-environments. (Individuals can be re-established as environments as well.) Nested environments are processed recursively, by adopting a finer spatial scale. Of course there's a limit to how fine this scale can be (given by the resolution of the eye). But there's another way to increase the scale of a location, which is simply to *approach* this location (a locomotion action). So approaching the object is another way of 'attentionally entering' it. (If the retinal scale becomes large enough, we will enter the object as a top-level environment, of course, which is the result we want.)
  - Connected to this—in the locomotion system, you can get from one environment *into* an adjacent environment. You end up with a bunch of links between environments, which can have as many cycles as you like. At this level, spatial LTM is just a directed graph of environments; we don't descend arbitrarily deep into recursion, but rather 'pop back' to some top level. (I think that *recognition* has something to do with that. When we get into a new environment, I suggest that *all* the LTM environments we aready know about are somehow fair game as candidates to be activated.
  - In the motor system, there is also in some sense a clearly fixed number of levels. There's an environment-centred motor system for locomotion. Then there are

---

[11] In the sense that each update operation is recursively defined.

various different 'agent-centred'—or more precisely, 'effector-centred'—motor systems for interacting with objects—for instance, for reaching with one hand, reaching with the other hand, reaching with both hands, etc. Each of these is in fact defined in a hybrid coordinate system, based on the relationship between the effector and the object to be interacted with. I suggest there are exactly two layers of hierarchy to these motor systems.[12] The highest layer is concerned with getting the effector *to* the object. (In reaching to grasp, this is the 'reach' component.) Another layer down is concerned with establishing the right kind of **contact configuration** with the object. (In reaching to grasp, this is the 'grasp' component.) This involves establishing the object as an environment, and then when a stable grasp is obtained, transiting *back* to a representation of the motor system as an environment.[13] There's obviously a limit to the size of objects we can manipulate, so even though any object can be recursively defined as an environment, there's no direct correlate for this recursion in the domain of manipulation. But there are some interesting types of recursion in the manipulation domain. Any object we interact with can be established as an agent in its own right; this is the basis for *causative* actions, in which we do a motor action whose effect is that some object 'does' *another* action. We can also use tools to manipulate small objects. I expect that the notion of causative actions will be the appropriate one to use to describe tool use.

- Functions
- Recognition
- Replay
- The picture that's emerging is one where there's recursive structure in how we *transition* from one SM state to the next, or from one LTM state to the next, but there are also certain natural 'stopping points' where we simply author information into a *flat* knowledge base, where all entities are represented in LTM in the same basic way. (As localist neural assemblies, I suppose.) I suggest that these stopping points are associated with the ends of right-branching XP sequences. These are places where LTM information is *authored.* (And, not coincidentally, where SM learning occurs, I expect.)

18.2 Domains

Environments, locations and individuals can exist in several different **domains**:

---

[12] At least, two that we need to worry about. I suspect that the reflex loops in the spinal cord implement the same kind of circuit, but at a level below that accessible to language.

[13] I need to work this out, but the basic idea is that everything gets reversed at this point.

- In the **spatial** domain, we construe environments, locations and individuals as opportunities to gather information about the world (passively, in a way which doesn't affect it).
    - The things we deploy our attention to are *physical locations*.
    - Each physical location can be established first as an object, with a property complex, and then afterwards as an environment. There are two different kinds of environment we can establish. One is as a semantic environment (where the object's individual properties compete to be 'semantically' attended to); the other is a spatial environment (which contains candidate locations and objects of its own).
    - The initial actions we can do are actions of physical attention, which involve 'directions of the body' (either to direct visual attention on a particular point, or to *touch* a particular point),[14] and actions of locomotion through our environment.
    -
- In the **temporal** domain, we construe environments, locations and individuals as opportunities to act in the world, and make changes to it.
    - The things we deploy our attention to are *situations*.
    - Each situation can be established first as a complex of 'situation types', and then afterwards as an environment, which makes available various substantive actions.
    - The actions we can do are 'substantive actions', which affect objects in some way (change their position or shape or both). These objects can be external objects (in which case we're minimally doing 'transitive' actions), but they can also be ourselves—so locomotion also counts as a substantive action of the agent on himself, and actions of physical attention, which require changes in body position, can also be parsed as substantive actions of the agent on himself. (Conversely, substantive actions on objects, e.g. 'grab', can also be interpreted as attention-directing actions, because they provide haptic information about the objects in question.)
    -
- In the **semantic** domain, we construe environments, locations and individuals as opportunities to learn about the *regularities* in the world, expressed as associations between concepts
    - The things we deploy our attention to are *LTM units*, as entities in their own right. (Or groups of

---

[14] 'Focal attention' probably relates to the 'directed' nature of these actions.

  - The actions we can do involve following associative connections between LTM units to get into a 'new LTM state'. (I suggest that these 'trains of thought' are implicated in the processes involved in relearning/consolidating hippocampal memories in cortex.)
  - We can choose to attend to the properties of this LTM unit (a semantic environment where individual properties compete), or we can establish the LTM unit as an environment, whereby it activates other LTM units which are 'associated' with it—i.e. which it *has*. This can be a physical possession environment—e.g. activating the LTM individual 'John' as a possession environment causes the set of LTM individuals which John *has* to be activated as candidate units. But it can also be an 'episodic memory' environment, activating all the properties which accrue to John as consequent states of episodes he was a participant in. (Which explains the past participle *has*.)

18.3 Motor systems

- The **current motor state** of a motor system is a distributed representation, where (to be simplistic) there is one unit for each degree of freedom—i.e. for each separately controllable joint. Actually, this is a simplification. Current motor state is partly computed from an internal forward model, that takes into account some notion of 'motor context', which represents current velocity (which isn't something that can be directly sensed[15]). I represent this forward model as a simple recurrent network (SRN), which takes an efferent copy of the most recent *motor command*, plus the current 'motor context', plus sensory inputs, and delivers a new motor context. The idea should be that this function is parameterisable, to take into account the fact that the dynamics of the plant can vary, e.g. depending on load, fatigue, orientation (this is Wolpert's idea in MOSAIC). You need to find the right *point* in the context layer space—the point that progresses the context layer through the right trajectory for the current plant dynamics.

- A **goal motor state** is represented using the same coordinate system—one unit for each degree of freedom. (Therefore there's a 1:1 connection between these banks of units.) Goal motor states are learned when the right kind of tactile stimulus is obtained—e.g. touch for the reach system, and a stable grasp for the grasp system. At this point, we learn that the goal motor state just *is* the current motor state.

- The simplest type of control is feedback control, where we just subtract the goal motor state from the current motor state. But again, we actually want something a little removed from this, because this simple scheme doesn't always work. What we need is a **control function**, that takes the current motor state and the goal motor state and generates a 'new' goal motor state, which is recurrently the new

---

[15] except through vision?

input to the control function (as well as providing the value to be subtracted from the current motor state to deliver the next motor impulse). The idea is that this can be understood as generating a trajectory of a 'virtual target' which you reach for directly using simple feedback control. The control function is also implemented as an SRN, where the context layer is the 'goal motor state'. The SRN can learn multiple alternative trajectories.

- My idea is that an open-class verb tells us about a way of initialising the motor controller SRN's recurrent layer,[16] so that it takes the target object on a *particular* 'virtual trajectory'. The connections in the SRN should be trained when the appropriate tactile stimulus is evoked. Training involves back-propagation through time. (Note that the object attended to at AgrP already gives us the target object as an 'actual' goal motor state. So we're already 'reaching for' the object at AgrP, in some sense. The open-class verb gives extra information about how we should perturb the 'default' goal motor state along various trajectories to achieve particular effects on the target object.)

18.4 Representing the results of motor actions: situation updates

- There's a connection between actions and situations. I suggest that the connection is like this: whenever we get a reward, we know whether it's a known kind of reward, or a new sort. (This has to do with dopamine, which responds to *unexpected* rewards.) I suggest that if the reward is a new sort, **we coin a new situation unit** to represent this new state. (Which is like creating a new LTM individual.) I suggest that when we do back-propagation through time, *we use this brand new unit as the biasing input*, so we begin to learn a *new* trajectory within the motor controller SRN, that's customised towards this type of input.[17] On the other hand, if the reward is expected, it's because a *known* potential reward situation has been identified, and already selected as the current bias on the motor controller. Thus back-propagation through time is a matter of refining, or adding to, an existing trajectory in the system.[18]

- The view which emerges from the above ideas is that open-class motor verbs are represented in two ways. On the one hand, within the LTM system, they are situations:

---

[16] or perhaps of *constantly biasing* the whole SRN.

[17] We also have to index this new unit to the previous situation, so that it's available the next time we encounter that situation.

[18] This idea relates very nicely to Schultz *et al.*'s (1997) model of the role of dopamine's signalling of expected rewards, and to Braver and Cohen's model of how control strategies are learned. (Maybe I should be doing temporal difference learning, rather than backprop through time–but I like the idea of *trajectories* you get in a recurrent network.)

localist LTM units.[19] (As such, they can be associated as individuals with all sorts of things: e.g. with candidate LTM individuals, with other candidate situations which are likely to be possible 'in' this type of situation, etc etc.) On the other hand, within the motor system they are specific trajectories within a motor controller. The connection between these things is as described in the previous bullet point: axiomatically, situations are rewards, and the system is set up to try and achieve rewards. When it expects a reward, it behaves appropriately. If the expected reward is forthcoming, the behaviour which led up to it should be reinforced (internally, as a motor programme, and also by relating this motor programme 'as a semantic unit' to the situation in which it was produced). If the agent gets an unexpected reward, it needs to treat the behaviour that generated that reward in a special way, and create a new LTM unit which ('sparsely') represents the reward state, so it can start to learn a brand new action in its repertoire (on the motor end) and a corresponding semantic unit which can be associated with the situation this new action occurred in. (Note: we will also try and learn about the *sensory cues* in the situation which allow us to predict the reward. So we will be learning a function which maps the sensory properties of the action's context situation onto the newly minted action.)

- A very good way to think about this is in terms of Braver and Cohen's 2000 model of PFC. The idea in their model is that when you get an unexpected reward, you gate open a function which maps sensory information onto activity in PFC (a function which is not otherwise allowed to operate). To start with, this function generates random (but probably sparse!) activity in PFC: something like a 'new unit'. I suggest that a reward situation independently has the effect of triggering learning of the motor controller trajectory which led up to the reward state. The point is that this learning is done in the presence of the new PFC state. We also assume that the LTM rep of the previous situation is made to link to the new PFC state, so that when you get to that previous situation next time you are (more) likely to 'adopt' the newly created PFC state as the new 'current SM situation'. If this does happen, we'll do the same action we did last time—i.e. activate the controller trajectory which the new PFC unit is associated with.

- Note that when a reward is unexpected, it comes at the *end* of the sensorimotor sequence which generates it (which in turn is done in a certain LTM context). In that case, the 'new situation unit' which fires at the reward point occurs after the motor action (/SM sequence) which generated the reward—and axiomatically represents this sequence. When a reward is expected, the situation is different: the LTM context activates a 'candidate situation' unit simply through associative connections, *in advance* of any SM sequence happening. The association is strong enough that this unit is selected as the next 'current SM state' unit (i.e. an SM update hap-

---

[19] Or more accurately, situation *types*. A genuinely individual situation is a unique moment in time, which is associated with a *complex* of situation types—see discussion in Section 13.9.1. But situation types are also individuals, represented in a localist way, I think.

pens).[20] At this point, a SM sequence occurs, step by step. If each step goes as expected, we end up in the unconditioned reward state. At this point, there is a *small* difference between the actual state reached and the state we thought we would be in—and we *tweak* the associations between the selected 'current situation' unit and (i) the context; (ii) the SM actions we actually carried out (which might differ slightly from those we planned). [I expect that language tells us about the 'tweaking' moments rather than the 'new creation' moments—except perhaps for DPs. The situations we get into are never completely new; that's just for babies. They have new elements, which present exploratory challenges. The actions we do are also not new ones, though we put them together in different ways.]

18.5 Objects and situation updates

- Following on from this: note I have to say that *objects* are situations too. (The object DP is the rightmost element in a transitive clause.) This is somewhat counterintuitive, but we can understand it in the light of the above discussion as follows. Firstly, objects are associated axiomatically with reward, since they are things we can pick up, hold, manipulate (all of which are special 'stable grasp' states which generate particular rewards). Secondly, note that objects are perceivable (e.g. visible). If an object is perceivable, our sensory environment contains cues to the tactile rewards available from touching/manipulating it—and these cues *also* put us into the right cognitive state to actually *carry out* the right kind of touching/manipulating action.[21] When we touch an object:
    - We create a PFC unit that in some sense tells us about the tactile niceness of the object. (That's the 'new situation'.) I suggest this is what's involved in 'making our current arm motor state the current *goal* arm motor state'.
    - We learn a function which maps our current visual state (and perhaps our current gaze-to-body angle too!) to this new current goal arm motor state. (The current goal arm motor state is obviously the *training signal* for this function, so that after training, the function can generate it for itself. (There's a clear difference between the *goal* motor state which is output from the visual function, and the *actual* motor state.)[22]

---

[20] Interesting idea: an SM update is not the same as an LTM update! One possibility is that sentences/LF structures only encode SM updates which are 'successful enough' to trigger LTM updates.

[21] Obviously tactile stimuli are only primary reinforcers—as adults, we get no kicks from touching arbitrary objects! So I'm thinking first about infants.

[22] A nice idea—the function doesn't map to the *current* PFC state—it's not allowed to do that. It maps to a *candidate* situation. (Of course, it'll actually map to *several* candidate situations.) If the aggregate activity in the candidate layer is high enough—or associated with enough dopamine, on aggregate—*then* we allow a new current situation to be selected. (*This* is what counts as a 'PFC update' in Braver and Cohen's model.) (I reckon PFC probably holds the candidate situations as well as the current situation.)

I suggest that the first of these steps ('make the current state the goal state!') is basically the same as the operation of 'creating a new current LTM situation unit' described above. But note that this 'situation' is only one *component* of a stable grasp situation. The arm is doing the right thing, in some sense; but the fingers aren't. It could be better! Because there's more reward to be had. When we grasp a cup, we do a *sequence* of things, which are nested inside one another. (Attend to agent, attend to cup, grasp.) The tactile state doesn't just cause the most recent one of these to be reinforced.

- I suggest that the second step above is what does the initial training of the visual classification system. Not exactly sure how, though. I can see how it trains 'affordance-based' visual object representations in parietal cortex. (Especially the useful saliency map/'location' representations.) But not how it enables learning of arbitrary semantic categories in inferotemporal cortex. That must involve more complex associations with reward, I guess. (The fact that objects can be used in certain motor programmes to achieve higher goals.)

- Note the PFC unit described above only tells us about the *actual touch moment* if the touch sensation is *unexpected*. For an adult, *all* touch sensations associated with objects are entirely expected, so all the PFC units activate purely on visual cues. (Whe we *actually* grab an object, we'll get exactly the touch sensations we predict. Which sounds related to DP-movement in a trans. clause.) At this point it seems like we don't actually need to touch every object we see any more. We switch visual attention from one object to another, looking for good opportunities to activate routines which involve particular *kinds* of touches on particular *kinds* of object. I suggest that a transitive cup-grabbing sentence tells us about a scenario where we have found one of these routines: having noticed a cup, we find a good match with the 'grab' routine, which is the best way we can find to achieve rewards in the current situation. Because there's an actual reward, the sentence is 'relevant' somehow. (I suggest that relevant sentences describe genuine reward situations. Not only do we spot that a reward situation is 'worth trying', and therefore execute a SM sequence; we also find that this sequence does indeed have its intended result. *That's* what's worth (i) doing some SM learning about; (ii) registering in LTM—which is a very similar thing—and hence that's the kind of thing that a sentence can describe.)

18.6 A note about coordinate systems

- We have saliency maps in several different coordinate systems: environment-centred, retina-centred, head-centred, body-centred. (And then all the motor maps.)

- Each of these maps has two layers: a layer of **current locations** and a layer of **candidate locations**.

- The ‘candidate locations’ maps represent ‘the current environment’ in some sense. The ‘current location’ map represents a focally attended location.

- I suggest that there are transformation functions which map bidirectionally between maps, using information about the observer’s current attentional state, thus:
    - A pair of functions map between the environment-centred and retina-centred saliency maps, using the environment-centred **current observer position** and **current observer gaze direction**. (One function maps between the ‘candidate locations’ maps and another very similar one maps between the ‘current locations’ maps.)
    - A similar pair of functions map between the retina-centred and head-centred saliency maps, using **current eye position** (in relation to head).
    - A similar pair of functions map between the head-centred and body-centred saliency maps, using **current head position** (in relation to body).
    - Similar pairs of functions map between the body-centred map and the various different motor maps, using **current motor state** in each system).

- The transformation functions are learned by ‘slow feature learning’: axiomatically, the invariant representations should not tend to change from one moment to the next, when the observer executes an arbitrary **attentional action** (e.g. an eye/head movement, a body rotation or a translation in the environment). The WTA links within each pair of maps help this learning. The difference between two successive invariant maps is treated as an error term, which the function learns to minimise.

- Once the transformation functions are good enough, this difference is also used to abductively infer the nature of ‘attentional actions’ like eye movements. For a given attentional action, we consult the saliency map which is supposed to be stable over this kind of action. If there is a difference between our predictions about this map and the actual map, we can infer that our attentional action was actually different from the one we thought we made. This is the origin of some of the plasticity in the sensorimotor system.

- Important idea: *selection of a current location* is the thing that allows us to activate a single LTM individual (i.e. to update the ‘current LTM individual’). Just making ‘attentional actions’ like eye/head/body movements doesn’t select a location. It just biases our competition towards certain *parts* of our current environment. Selection of a current location is completely different. It can be initiated by bottom-up saliency, or by a strongly-active *candidate* LTM individual. In order to **establish** an individual, we need to find a ‘real’ salient retinal location, and we need to activate a ‘sufficiently good’ property complex in IT. When that happens, we either trigger an existing LTM individual whose location and properties are a close enough match (‘recognition’), or if this doesn’t happen, we activate a *new* LTM individual. In either case, we have an update in the current LTM individual. (Note that this update only occurs after some SM experience. A winning SM location will have been selected. Then a spat. frequency will have been selected, and a property complex will have been activated. This sequence will be *facilitated* by changes in the pattern of activity in the LTM units, which can often start off representing a bunch of overlaid attentional plans, and will end up strongly suggesting one particular plan (if we're about to recognise a particular LTM individual). My idea is that once we have 'established' a new LTM individual, there's a process whereby the (stored) SM operations we have taken to bring this about are written into LTM (a process which involves bringing about a sequence of LTM updates or transitions).

- Having selected a current location, there are quite likely to be attentional actions which respond to this. For instance, if we select a peripheral retinal location as the current location, we will (*probably*) make an eye movement to foveate this location, and we may then make a head movement to centre the eye in the head, and we may then make a body movement, to have the body pointing in the same direction as the head. But we don't have to do these things.

18.7 DPs and semantic memory

- In semantic memory, there is also the notion of an environment. A pure **semantic environment** is a state of LTM where activating a property assembly is sufficient to activate an individual. This individual will be the individual which has that property assembly. Of course, it could be a *group* individual. My idea is that the account of quantification I give in Chapter 12.6) is an account of this kind of environment.

- Agents can also function as semantic environments. The semantic environment of an agent is the set of objects it *possesses*.

- There are also hybrid environments. For instance, a location in the world has physical contents, but also semantic associations. If I'm representing a particular 'current location' as my current environment, and there's a particular (token) dog associated with this environment, then the semantic activation of a generic 'dog' category will (or may) be enough to trigger activation of the token dog. Certainly that's the path of least resistance. Of course, if I'm experiencing the world, I will evoke a dog *property complex*—something much richer than a simple dog prototype. If I do this, there are various possibilities. If the property complex matches that of the unique dog associated with this environment, then the category 'dog' has the expected effect, in retrieving the expected LTM individual (showing that the semantic associations of the environment are working as expected). I suggest that this underlies the 'LTM' concept of definiteness: we would describe this scenario using the DP *THE dog*. (In other words: definiteness is all about the LTM and perceptual machinery supporting *object recognition*.) But in other situations, our expectations aren't matched, and we attend to a dog whose property complex does not match that of the dog associated with the environment. In this case, we must activate a different LTM dog individual instead. This might be an existing individual, or a new one. I suggest that the determiner *a* describes the case where it's new. If it's a known individual, I suggest there's a mismatch between our physical and environment representations: we need to tweak the semantic associations of the physical environment. I suggest that possessive determiners describe this process. For instance, say we notice *John's dog* in my house, where it's not expected. This is a surprise, and as a consequence, we need to associate my house with the semantic environment created by activating 'John' (the LTM individual).

- The notion of 'recognition' just mentioned is analogous to the 'triggering of a new situation' in Section 18.1. The idea is that activation of a 'new current LTM individual' (whether its an existing one or a newly-created one) is due on the one hand to a sensorimotor (attentional) sequence, which steps through various stages and ends up evoking a particular sensory manifold, and on the other to transitive associations within the system of *candidate units*, which link *forward* from the current LTM state but also *backward* from the new manifold. The set of units which become active will do so because they mutually support one another. My idea is that once they become active, we enter a new phase where the active sensorimotor units support a step-by-step update of the current LTM state, until a new state is created. The operation of activating the new LTM individual is the pivot between the two phases. I suggest that the sensory manifold evoked at the 'end' of the sensorimotor sequence (which will include a location, relative spatial frequency and property complex) is somehow *one component of* the state in which the new current LTM individual activates. (Maybe the relationship between these two components is captured by the relationship between a 'syntactic position' and the XP which 'occupies' that position.)[23]

- The tracking processes that happen during sensorimotor experience provide a bridge between the 'current' and 'next' LTM states (which are *discrete* in nature—there's a discrete update operation which takes us from one LTM state to the next). LTM individuals are associated with *tracked* locations, which can change their values during the course of sensorimotor experience. At the start of SM experience, the binding of a location to an LTM individual is done through 'recognition'—i.e. through a combination of SM state (bottom-up) and transitive associations within the 'current units' layers (top-down). At the end of SM experience—probably within the DP system—a *new* static location is associated with the *same* LTM individual, in the *updated*

---

[23] Another aspect of the 'pivot' is the set of 'current' SM units. I think these represent 'sensorimotor working memory'—something like 'the current WM episode'. As I discuss in Chapter 3, a WM episode can be *forward-looking*, in the sense that it generates a sequence of SM actions, in the manner of a plan, but also *backward-looking*, in that it retains a representation of a SM sequence that has just occured. Its backward-looking nature allows it to serve as the representation which supports simulated replay of SM sequences.

situation/environment. This ensures the spatiotemporal continuity of objects, as we conceive them. (This idea relates pretty well to the model of the relationship between object and episode representations in Section 11.3.)

- The fact that the subject is tracked, together with the fact that XPs describe a replayed SM experience *after tracking is complete*, means that the *SM* component of IP-max (the Wiskott function, which evokes the environment-centred location of the subject) is able to tell us about the *new* location of the tracked subject immediately. (The location at which the subject was *recognised* at the start of the SM experience, and associated with an LTM individual, is no longer available. But the identity of the LTM individual is still available.)

18.8 Non-SM operation and pseudorehearsal

I like the idea that we can also get something analogous to 'SM experience' just by putting LTM into a random state and seeing what update the LTM network finds for itself. I suggest we can learn from this kind of 'experience' in just the same way, by making tweaks to the network's weights. This is a neat paradigm for pseudorehearsal.

18.9 Interpretation of X-bar syntax

- The 'current discourse context' is...
- XP—-the maximal projection—is...
- The X head is...
- The XP max. projection tells us about (i) a **sensory medium**; and (ii) a **function**, which maps representations in this medium onto some more derived *LTM* representation. (e.g. IP tells us about the raw visual manifold, but also about the Wiskott function, which maps this manifold onto a point in an environment-centred coordinate system.) I suggest that the [Comp,X] of this XP *continues* to denote this function.
- I suggest that the YP which *occupies* [Comp,X] position denotes a function which takes its own source of raw sensory data onto a sensorimotor representation in a coordinate system specific to YP.
- I suggest that the 'slot-filler' relationship (something like binding) between [Comp,X] position and the YP which occupies this position denotes a *separate* **transformation function** which takes the *output* of the function denoted by XP, and returns values which are somehow constrained to be the same as / consistent with the values returned by the YP function. (This constraint drives the *learning* of the transformation function.)

- ‘[Spec,XP]’ (the *position*) is...
- ‘[Comp,X]’ (the *position*) is...
- The YP that *occupies* [Spec,XP] is...
- The ZP that *occupies* [Comp, X] is...
- A complete (single) XP structure can be understood as saying something both about an *individual*, and about that individual’s *environment*.
- Every XP tells us about an operation which gets us into a particular **motor environment**.
- Following on from this: different *types* of head contribute different sorts of information in their own right.
  - IP says ‘we’re in the top-level environment’.
  - VP says ‘use the current motor environment’. (A particular V head just says what ‘mode’ it’s used in.)
  - AgrP says ‘refine the current motor environment, by choosing a focal SM environment’.
  - PP says (first approximation): ‘we’re establishing the current object as an environment’.
  - DP says (first approximation): ‘here’s what’s going on in the LTM object representation system’.
- In a right-branching sequence of XPs, there’s some idea that the first XP starts a *tracking* operation, which persists through the operations described by the other XPs. The tracking operation *drives* changes to the observer’s motor system—for instance, smooth tracking eye/head movements—and it *survives* other changes—e.g. the locomotion or rotation of the observer (or observed objects) in the environment. I suggest that the ‘tracking’ motor movements create transitory motor states which provide an additional input to the ‘transformation function’ for this XP (described above).
- A domain of head movement is..
- There’s some connection between the top and the bottom of a right-branching chain of XPs.
- I suggest that the top—and bottom—of a right-branching chain of XPs somehow both tell us about (i) the activation of a new LTM unit through ‘recognition’, which will eventually lead to a new ‘sparse’ ‘fact’ being written to hippocampal/temporal LTM; and (ii) a piece of SM learning being done.

- Case-assigning positions are...

- Movement to Case-assigning positions describes the following constraint in the network:...

- The tweaks made by sensorimotor learning are manifest in syntax in the following constraint:...

- The possibility of (left-)recursion in syntax reflects the following fact about the network:...

- In an IP describing an episode there's a right-branching sequence of XPs, each of which has a DP at its Specifier. We're moving from one action-related context to another. At each point, we *also* deploy focal attention.

- I quite like the idea that [Spec,XP] tells us about something that happens 'in the middle' of a discrete update of LTM. Getting from one 'stable' LTM situation to another involves going through a few 'intermediate operations' (see above), reaching states *which are not stable in themselves.* I like the idea that these intermediate stages *do* in fact describe stable, complete updates *in some other domain.* Thus, for instance, attending to and establishing an object (an agent or a target) doesn't in itself update *the situation.* It gets us 'part-way' towards such an update. However, in the *stative* system, which registers LTM individuals and their location / semantic associations with the current environment, establishing an object *is* a complete update operation. (It has its own incomplete stages, which are represented by its own right-branching XP structure.)[24]

---

[24] And perhaps some of these stages can in turn be complete 'proposition-like' things—which would introduce the possibility of embedded sentences.

Chapter 19

Relationship of the theory to existing work

If the ideas described in this work are on the right track, then there are clear consequences both for research in the Minimalist tradition, and for research in sensorimotor modelling. Put simply, if there is a relationship between the theories developed in these two different areas, then research in one area can be used to inform research in the other. The discovery of new aspects of the sensorimotor system can drive the generation of hypotheses about corresponding elements of a Minimalist syntactic model. Conversely, the postulation of a new syntactic projection, on the strength of linguistic argumentation, can drive the formulation of hypotheses about the sensorimotor role of the hypothesised projection. Arguments can begin to draw equally on linguistic and sensorimotor data, and components of the model which can be motivated from both types of data will be strongly supported. In fact, the arguments developed in Chapter **??** have precisely this form. It is thus to be hoped that a rapprochement will occur between theoretical linguists and sensorimotor modellers, with researchers gaining knowledge in both of these fields.

There are also implications of the present work to other approaches to language study. Most directly, there are many other theorists who have proposed a relationship between linguistic structure and cognitive structure. In Section 19.1, I discuss the main proposals in this area, and the relationship of these proposals to the current work. In Section 19.2, I consider the relation of the current work to work in constructivist models of syntax. Section 19.3 considers work in sentence processing models (both parsing and generation). Section 19.5 discusses issues related to the localisation of language-specific processes in the brain; Section 19.6 discusses relations with research into language development, and Section 19.8 discusses relations with research in the formal semantics of natural language.

19.1 Related work in cognitive linguistics

19.1.1 The Arbib-Rizzolatti model

Rizzolatti and Arbib (1998), and Arbib (2005). The focus on mirror neurons is a little different in these papers.

For me, the importance of mirror neurons is mainly in allowing a mapping between heard words and spoken words (c.f. Section **??**).

19.1.2 Ullman's model

Ullman (2004) is very relevant. Ullman has compiled a detailed catalogue of evidence from neuroimaging and disorder studies to suggest that the neural mechanisms underlying grammar are related to those underlying procedural memory, and that the neural mechanisms underlying the lexicon are related to those underlying declarative memory. I have focussed more on grammar than on the lexicon in my model, but in each case, these are exactly the predictions which I would make.

Ullman's declarative/procedural model paints quite a broad-brush picture of a disctinction between lexical and syntactic knowledge/processing. My model can be thought of as providing additional detail to the model, particularly as regards syntactic processing, but it is certainly very compatible with it.

19.1.3 Hurford's model

Hurford (2003) is about correspondences between cognitive processes in perception and natural language semantics. But Hurford doesn't want to postulate similar correspondences relating to syntax. My model, on the other hand, looks for a sensorimotor basis for natural language syntax as well as semantics.

19.1.4 Dominey's model

There are a number of similarities between the model presented here and that given by Dominey (1997; see also Dominey *et al.*, 2003). (. . .)

19.1.5 Calvin and Bickerton's model

Calvin and Bickerton (2000). This is the only model which seeks to give an interpretation of specific neural mechanisms in terms of transformational grammar. Indeed, Calvin and Bickerton's syntactic model is a simplified version of Minimalism, just as mine is.

19.1.6 Corballis' model

Corballis (2002).

19.1.7 Feldman/Narayanan's models

The L_0/NTL task is very similar. The notion of 'executable' models as semantic representations is the same. Their syntactic models are quite different.

19.1.8 Cognitive linguistics models

Osgood (1971), Givón (2002) and others.

19.2 Related work in empiricist approaches to syntax

Here include all the work on constructivism and empiricist approaches to syntactic theory.

19.2.1 Recurrent networks in sentence processing

Firstly, people working within the simple recurrent network paradigm, who analyse sentences as Markov processes: Elman (1990; 1995), Christiansen and Chater (1999) and so on.

19.2.2 Construction grammars

Secondly, people working with variants on construction-based grammars: Goldberg (1995), Jackendoff (2002), Tomasello (2003).

19.2.3 Statistical linguistics

Thirdly, mention current computational work in statistical natural language processing: Abney (1996). Here again, there is work exploring pure Markovian models of sentence structure (especially in part-of-speech tagging and n-gram-based models of lexical semantics) as well as in hybrid symbolic/statistical models: see e.g. Collins (1996), and much subsequent work.

19.2.4 The principles-and-parameters paradigm

Finally, we must not forget that traditional generative grammar foresees a role for empirical tuning of the constraints provided by universal grammar. As is well known, the Chomskyan model of language is one of 'principles and parameters', where principles are constraints on the possible structure of languages which are specified biologically, and parameters are constraints which are established through exposure to the speech of others in the learner's language community. The main difference between generative grammar and work in constructivist linguistics is one of emphasis: Chomskyan linguists have traditionally been concerned with identifying parameters as theorists, rather than with the learning architectures which would allow them to be identified automatically.

19.2.5 Summary

In summary, there is a good consensus that any grammar needs to include reference to a statistical mechanism for learning regularities in the input speech stream. The more extreme empiricist theories hold that knowledge of language is acquired by a very general-purpose statistical learning engine, so that the structures in mature language are due almost entirely to statistical regularities in the language a child is exposed to. Less extreme views suggest that a core universal set of language-specific principles are tuned by a statistical component which takes the speech stream as input. The theory I have presented is of this latter type. It includes some elements which are present at birth (though many of the mechanisms which in generative grammar are assumed to be language-specific are in my model given a general sensorimotor role). But it also includes a statistical component which can be expected to deliver the same kinds of collocation-based structures as SRNs, construction grammars, or statistical context-free grammars.

Also mention a link to people working on theories of the cultural evolution of language: e.g. Kirby. The empirically learned component of any language is subject to cultural evolution, in the kind of ways described by Kirby and others.

19.3 Related work in sentence parsing and generation

Here include more on incremental interpretation. See review by Crocker. Talk about Spivey *et al.* (2002)—this fits nicely with the idea that interpreting a DP involves accessing the object file associated with that DP. For generation, refer to Bock and Griffin, who find attention to an object just before the DP referring to it is generated; the action of attention is assumed to help prepare the referring expression. Note that there's nothing strictly analogous to these eye movements when the picture is actually observed. I think we have to assume that during sentence generation, eye movements are controlled by the same circuits that control verbal activity, rather than by the regular 'action understanding' circuitry, which is slightly different.

Also mention in passing dynamic syntax: Kempson and Gabbay.

19.4 Related work in neuroscience

19.4.1 Damasio's account of second-order narratives

The model introduced in Damasio (1999) has quite a lot in common with the model I'm putting forward. Damasio is very big on the idea that sentences report cognitive episodes, and that these episodes are extended in time, and are composed of sequences of mental events and processes.

19.5 Related work in brain localisation

I have mentioned various possibilities for the localisation of elements of the sensorimotor model developed here (see e.g. Sections **??** and **??**). However, we have not considered the important question of whereabouts in the brain the language-specific components of the model are found.

In general, my model predicts that there will be two language-related areas: one for storing the mental lexicon, and one for linking word sequences to sensorimotor sequences. The hypothesis that language involves these two separate mechanisms is as old as the hills; traditionally, a frontal region called **Broca's area** is associated with syntactic knowledge and processing, while a region in the left superior temporal gyrus called **Wernicke's area** is associated with knowledge of the semantics of individual words. Damage to these two areas is traditionally held to result in selective impairment of syntax and semantics respectively; see e.g. Damasio and Damasio, 1992; Pinker (1994 for recent versions of this claim. In recent years an alternative hypothesis has arisen, contending that words and syntactic rules emerge from the operations of a single neural structure (see e.g. Bates and MacWhinney, 1989; Seidenberg, 1997). While the debate about the role of Broca's and Wernicke's areas is far from resolved, it is clear that my model adopts a fairly traditional view of these two areas. There is certainly plenty of recent evidence to support this traditional view; see for instance Ullman *et al.* (2005).

Saygin *et al.* (2004) have found that aphasic patients tend to have deficits in non-linguistic action comprehension, but that the severity of the aphasia and of the action comprehension deficit are not strongly correlated…

19.5.1 The mental lexicon and left temporal-parietal structures

19.5.2 Broca's area and sensorimotor sequencing

Cite Greenfield (1991) (cited in Dominey *et al* (2003)) for evidence that in children under 2, Broca's area represents sensorimotor as well as linguistic sequencing.. Lots of other citations from Dominey needed here too.

Cite Müller and Basho (2004) for the idea that left prefrontal cortex is not just specialised for language, but is also activated by sensorimotor and working memory processes. Müller and Basho propose that during language acquisition, the left PFC learns mappings between these nonlinguistic representations and linguistic ones. (I think.)

19.5.3 The cerebellum, motor learning and inflectional morphology

It is well known that the cerebellum is involved in motor (or 'procedural') learning (see e.g. Molinari *et al*, 1997). In the last ten years or so, it has been found that the cerebellum is also involved in a range of more cognitive functions, particularly to do with language and phonological short-term memory (Silveri and Misciagna, 2000). Cerebellar damage appears to cause certain fairly mild but characteristic linguistic deficits, in particular a variety of agrammatism in which subjects tend to leave off verb inflections when generating sentences (Silveri *et al.*, 1994; Justus, 2004) and a failure to diagnose ill-formed sentences due to agreement mismatches (Justus, 2004). In my sensorimotor model of syntactic representations, the process of generating verbal inflections is related to the process of motor learning, generating fairly bold predictions about co-occurrences of deficits in these two apparently unrelated abilities. The recent findings which suggest that these two processes share a neural substrate are therefore quite interesting.

19.6 Related work in developmental linguistics

Include discussion of the role of joint attention, and the importance of recognising intentionality: Baldwin, Tomasello.

19.7 Related work in models of memory

Recall that the next-state function of the sensorimotor sequencing model is intended to be an abstract description of episodic memory formation in the hippocampus. The basic idea is that the hippocampus stores memories of (salient) sensorimotor sequences. The theory of hippocampal function which I am assuming is that of Lisman and Otmakhova (2001), in which the hippocampus uses theta cycle phase precession to store information about the expected sequence of sensory states.

The closest model to the account of episodic memory given in this book is Shastri's model of hippocampal function (Shastri, 2001; 2002. This model also sees semantic representations of sentences stored in hippocampal episodic memory traces, and it also makes use of the theta cycle. The main difference is that in Shastri's model, there is an explicit representation of thematic roles: at each individual phase in the theta cycle, a 'thematic role unit' such as 'agent', 'patient' etc is active, concurrently with the cortical representation of the object which has the associated thematic role. For instance, to represent the semantics of *The man chased the dog*, 'agent' might fire with 'man' at phase offset 1 and 'patient' might fire with 'dog' at phase offset 2, while 'chase' would fire at all phase offsets. Shastri's model thus offers a temporal synchrony solution to the problem of binding object representations to thematic roles. In my account, 'man' and 'dog' also fire at different offsets in the theta cycle, but their thematic roles are represented *implicitly*, by their position in the sensorimotor sequence; 'man' would fire first, and 'dog' would fire next, indicating that these two entities are agent and patient respectively. In both models, there is a notion of binding by temporal synchrony, but in my model, it is used to represent something which is *inherently* sequential, namely a sensorimotor sequence, while in Shastri's it is an essentially arbitrary solution to the problem of representing multiple role-value bindings in what is at heart a non-dynamic event representation. My model affords a reasonably clear story about how the hippocampus could evolve to encode event representations from an initial state in which (as in the rat) it merely encodes sequences of predicted sensory states. It is harder to understand how Shastri-style event representations could have emerged from such an initial state.

Shastri's model may encounter difficulties representing the semantics of sentences with nested relative clauses, given that the number of thematic role bindings is constrained by the number of phase offsets in the theta cycle (thought to be around 7). His model is also less transparently linked to a sensorimotor model of event perception. It is not clear how Shastri's explicitly declarative thematic role representations such as 'agent' and 'patient' are computed from perception, and there is not as yet any good evidence that the brain ever stores such declarative representations.

19.8 Related work in formal semantics

Here include reference to dynamic semantics: Heim (1982), Kamp and Reyle (1993). Also work in dynamic logic more generally: e.g. Muskens *et al.* (1997).

Bibliography

Abney, S. (1987). *The English Noun Phrase in its Sentential Aspect*. Ph.D. thesis, MIT, Cambridge, MA.

Abney, S. (1996). Statistical methods in linguistics. In J. Klavans and P. Resnick, editors, *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. MIT Press.

Abraham, W., Logan, B., Greenwood, J., and Dragunow, M. (2002). Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *Journal of Neuroscience*, **22**, 9626–9634.

Abrahams, S., Morris, R., Polkey, C., Jaroszc, J., Cox, T., Gravese, M., and Pickering, A. (1997). Hippocampal involvement in spatial and working memory: A structural MRI analysis of patients with unilateral mesial temporal lobe sclerosis. *Brain and Cognition*, **41**(1), 39–65.

Abrams, R. and Christ, S. (2003). Motion onset captures attention. *Psychological Science*, **14**(5), 427–432.

Adams, A. and Gathercole, S. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research*, **38**, 403–414.

Agre, P. and Chapman, D. (1987). Pengi: An implementation of a theory of activity. In *Proceedings of the American Association for Artificial Intelligence*, pages 268–272.

Aguirre, G. and D'Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain*, **122**, 1613–1628.

Alexander, G., DeLong, M., and Strick, P. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, **9**, 357–381.

Alexander, M., Naeser, M., and Palumbo, C. (1990). Broca's area aphasias: Aphasia after lesions including the frontal operculum. *Neurology*, **40**, 353–362.

Alexiadou, A. (1997). *Adverb placement: A case study in antisymmetric syntax*. John Benjamins, Amsterdam.

Alexiadou, A., Haegeman, L., and Stavrou, M. (2007). *Noun Phrase in the Generative Perspective*. Mouton de Gruyter, Berlin.

Allan, K., Wilding, E., and Rugg, M. (1998). Electrophysiological evidence for dissociable processes contributing to recollection. *Acta Psychologica*, **98**, 231–252.

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, **4**, 267–278.

Alvarez, G. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, **15**(3), 122–131.

Amodio, D. and Frith, C. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, **7**, 268–277.

Andersen, R. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London B*, **352**(1360), 1421–1428.

Anderson, J. and Lebiere, C. (1998). *The Atomic Components of Thought*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Anderson, J., Gilmore, R., Roper, S., Crosson, B., Bauer, R., Nadeau, S., Beversdorf, D., Cibula, J andd Rogish, M., Kortenkamp, S., Hughes, J., Gonzalez Rothi, L., and Heilman, K. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke-Geschwind model. *Brain and Language*, **70**, 1–12.

Anderson, M. (2010). Neural re-use as a fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, **33**(3), 245–313.

Arbib, M. (1981). Perceptual structures and distributed motor control. In V. Brooks, editor, *Handbook of Physiology: Section II The Nervous System. Part 1: Motor Control*. American Physiological Society.

Arbib, M. (1998). Schema theory. In A. Arbib, editor, *Handbook of Brain Theory and Neural Networks*, pages 993–998. MIT Press, Cambridge, MA.

Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, **28**(2), 105–167.

Arbib, M. (2006). Aphasia, apraxia and the evolution of the language-ready brain. *Aphasiology*, **20**, 1125–1155.

Arcizet, F., Jouffrais, C., and Girard, P. (2008). Natural textures classification in area V4 of the macaque monkey. *Experimental Brain Research*, **189**, 109–120.

Aron, A., Robbins, T., and Poldrack, R. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, **8**, 170–177.

Asaad, W., Rainer, G., and Miller, E. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, **84**, 451–459.

Auerbach, S., Allard, T., Naeser, M., Alexander, M., and Albert, M. (1982). Pure word deafness: Analysis of a case with bilateral lesions and a defect at the prephonemic level. *Brain*, **105**, 271–300.

Averbeck, B. and Lee, D. (2007). Prefrontal correlates of memory for sequences. *Journal of Neuroscience*, **27**(9), 2204–2211.

Averbeck, B., Chafee, M., Crowe, D., and Georgopoulos, A. (2002). Parallel processing of serial movements in prefrontal cortex. *PNAS*, **99**(20), 13172–13177.

Averbeck, B., Sohn, J., and Lee, D. (2006). Activity in prefrontal cortex during dynamic selection of action sequences. *Nature Neurosceience*, **9**(2), 276–282.

Axmacher, N., Elger, C., and Fell, J. (2008). Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain*, **131**(7), 1806–1817.

Baars, B., Motley, M., and MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, **14**, 382–391.

Babiloni, C., Del Percio, C., Babiloni, F., Carducci, F., Cincotti, F., Moretti, D., and Rossini, P. (2003). Transient human cortical responses during the observation of simple finger movements: A high-resolution EEG study. *Human Brain Mapping*, **20**(3), 148–157.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *TICS*, **4**(11), 417–423.

Baddeley, A. and Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, **129**(1), 126–145.

Baddeley, A. and Hitch, G. (1974). Working memory. In G. Bower, editor, *The psychology of Learning and Motivation*, pages 48–79. Academic Press.

Baddeley, A. and Warrington, E. (1970). Amnesia and the distinction between long-term and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **9**, 176–189.

Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, **105**(1), 158–173.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science*, **3**, 133–140.

Baillargeon, R. (1995). Physical reasoning in infancy. In M. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 179–186. Oxford University Press, Oxford, UK.

Baillargeon, R. (1998). Infants' understanding of the physical world. In M. Sabourin, F. Craik, and M. Robert, editors, *Current Directions in Psychological Science*, pages 503–529. Psychology Press, London.

Baillargeon, R. and Wang, S. (2002). Event categorization in infancy. *Trends in Cognitive Sciences*, **6**(2), 85–93.

Baker, C., Keysers, C., Jellema, T., Wicker, B., and Perrett, D. (2001). Neuronal representation of disappearing and hidden objects in temporal cortex of the macaque. *Experimental Brain Research*, **140**, 375–381.

Baker, M. (2008). *The Syntax of Agreement and Concord*. Cambridge University Press, Cambridge, UK.

Balan, P. and Ferrera, V. (2003). Effects of gaze shifts on maintenance of spatial memory in macaque frontal eye field. *Journal of Neuroscience*, **23**(13), 5446–5454.

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, **62**, 875–890.

Baldwin, D. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, **29**(5), 832–843.

Baldwin, D., Markman, E., Bill, B., Desjardins, R., Irwin, J., and Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, **67**, 3135–3153.

Ballard, D., Hayhoe, M., Pook, P., and Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, **20**(4), 723–767.

Bar, M. and Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, **38**(2), 347–358.

Bar, M., Tootell, R., Schacter, D., Greve, D., Sischl, B., Mendola, J., Rosen, B., and Dale, A. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, **29**(2), 529–535.

Barner, D., Wood, J., Hauser, M., and Carey, S. (2008). Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition*, **107**, 603–622.

Barone, P. and Joseph, J.-P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research*, **78**, 447–464.

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., Jeffery, K., and Burgess, N. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, **17**(1–2), 71–97.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, **22**, 577–660.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, **59**, 617–645.

Barsalou, L., Simmons, W., Barbey, A., and Wilson, C. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, **7**(2), 84–91.

Barth, H., Kanwisher, N., and Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, **86**, 201–221.

Bartlett, F. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, Cambridge, UK.

Bates, E. and MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney and E. Bates, editors, *The crosslinguistic study of sentence processing*, pages 3–73. Cambridge University Press, Cambridge, UK.

Battaglia-Mayer, A., Caminiti, R., Lacquaniti, F., and Zago, M. (2003). Multiple levels of representation of reaching in the parieto-frontal network. *Cerebral Cortex*, **13**, 1009–1022.

Bauer, W. (1997). *The Reed reference grammar of Māori*. Reed books, Auckland, NZ.

Behne, T., Carpenter, M., and Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, **8**(6), 492–499.

Behrens, C., van den Boom, L., de Hoz, L., Friedman, A., and Heinemann, U. (2005). Induction of sharp wave-ripple complexes in vitro and reorganization of hippocampal networks. *Nature Neuroscience*, **8**(11), 1560–1567.

Beiser, D. and Houk, J. (1998). Model of cortical-basal ganglionic processing: Encoding the serial order of sensory events. *Journal of Neurophysiology*, **79**(6), 3168–3188.

Beisteiner, R., Höllinger, P., Lindinger, G., Lang, W., and Berthoz, A. (1995). Mental representations of movements. Brain potentials associated with imagination of hand movements. *Electroencephalography and Clinical Neurophysiology*, **96**, 189–193.

Belletti, A. (1990). *Generalized Verb Movement*. Rosenberg and Sellier, Turin.

Belletti, A. (2001). Agreement projections. In M. Baltin and C. Collins, editors, *The Handbook of Contemporary Syntactic Theory*, pages 483–510. Blackwell, Oxford.

Bergen, B. and Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In J. O. Östman and M. Fried, editors, *Construction Grammar(s): Cognitive and Cross-Language Dimensions*, pages 147–190. John Benjamins, Amsterdam. Reprinted in V. Evans, B. Bergen and J. Zinken (eds.), The Cognitive Linguistics Reader. Equinox. 2007.

Bi, G. and Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, **18**, 10464–10472.

Bichot, N. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience*, **2**, 549–554.

Bichot, N., Rao, S., and Schall, J. (2001a). Continuous processing in macaque frontal cortex during visual search. *Neuropsychologia*, **39**, 972–982.

Bichot, N., Thompson, K., Rao, S., and Schall, J. (2001b). Reliability of macaque frontal eye field neurons signaling saccade targets during visual search. *Journal of Neuroscience*, **21**, 713–725.

Bickerton, D. (1981). *Roots of Language*. Karoma Publishers, Ann Arbor, MI.

Bierwisch, M. (1999). How much space gets into language? In P. Bloom, M. Peterson, L. Nadel, and M. Garrett, editors, *Language and Space*, pages 31–76. MIT Press, Cambridge, MA.

Bilkey, D. and Clearwater, J. (2005). The dynamic nature of spatial encoding in the hippocampus. *Behavioural Neuroscience*, **119**, 1533–1545.

Bittner, D., Dressler, W., and Kilani-Schoch, M. (2003). *Development of Verb Inflection in First Language Acquisition: A Cross-Linguistic Perspective*. Mouton de Gruyter, Berlin.

Blakemore, S., Wolpert, D., and Frith, C. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, **1**, 635–640.

Blakemore, S., Wolpert, D., and Frith, C. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences,*, **6**(6), 237–242.

Blakemore, S., Bristow, D., Bird, G., Frith, C., and Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision-touch synaesthesia. *Brain*, **128**(7), 1571–1583.

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith, editor, *The Handbook of Phonological Theory*, pages 206–244. Blackwell, Oxford.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press, Cambridge, MA.

Blumenfeld, R. and Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: An integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist*, **13**(3), 280–291.

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **12**(4), 575–586.

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin and Review*, **3**(4), 395–421.

Bock, K., Irwin, D., and Davidson, D. (2004). Putting first things first. In J. Henderson and F. Ferreira, editors, *The Interface of Language, Vision and Action: Eye Movements and the Visual World*, pages 249–278. Psychology Press, New York.

Bonda, E., Petrides, M., and Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, **16**(11), 3737–3744.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, **25**, 151–188.

Braine, M. (1976). Children's first word combinations. *Monographs of the society for research in child development*, **41**(1), 1–104.

Branigan, H. (2007). Syntactic priming. *Language and Linguistics Compass*, **1**(1–2), 1–16.

Brass, M. and Haggard, P. (2010). The hidden side of intentional action: the role of the anterior insular cortex. *Brain Structure and Function*, **214**, 603–610.

Braver, T. and Cohen, J. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell and J. Driver, editors, *Attention and Performance XVIII: Control of cognitive processes*, pages 713–737. MIT Press.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D., and Pylkkäanen, L. (in press). Syntactic structure building in the anterior temporal lobe during natural story listenings. *Brain and Language*.

Broadbent, D. (1985). A question of levels: Comments on McClelland and Rumelhart. *Journal of Experimental Psychology: General*, **114**(2), 189–192.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, **47**, 139–159.

Brookshire, R. (1997). *Introduction to Neurogenic Communication Disorders*. Mosby-Year Book, St. Louis.

Browman, C. and Goldstein, L. (1995). Dynamics and articulatory phonology. In R. Port and T. van Gelder, editors, *Mind as Motion: Explorations in the Dynamics of Cognition*, pages 175–193. MIT Press, Cambridge, MA.

Bruner, J. (1974). The ontogenesis of speech acts. *Journal of Child Language*, **2**, 1–19.

Brunet, E., Sarfati, Y., Hardy-Baylé, M., and Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage*, **11**, 157–166.

Buccino, G., Binkofski, F., Fink, G., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R., Zilles, K., Rizzolatti, G., and Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, **13**, 400–404.

Buckner, R. (1996). Beyond HERA: Contributions of specific prefrontal brain areas to long-term memory retrieval. *Psychonomic Bulletin and Review*, **3**, 149–158.

Buckner, R. (2003). Functional-anatomic correlates of control processes in memory. *Journal of Neuroscience*, **23**(10), 3999–4004.

Buckner, R. and Wheeler, M. (2001). The cognitive neuroscience of remembering. *Nature Reviews Neuroscience*, **2**, 624–634.

Buckner, R., Koutsaal, W., Schacter, D., Wagner, A., and Rosen, B. (1998). Functional-anatomic study of episodic retrieval using fMRI. I. Retrieval effort versus retrieval success. *Neuroimage*, **7**, 151–162.

Bullock, D., Cisek, P., and Grossberg, S. (1998). Cortical networks for control of voluntary arm movements under variable force conditions. *Cerebral Cortex*, **8**, 48–62.

Burgess, N. and Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, **106**, 551–581.

Burgess, N. and Hitch, G. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences*, **9**(11), 535–541.

Burgess, N. and O'Keefe, J. (1996). Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus*, **6**(6), 749–762.

Burgess, N., Jackson, A., Hartley, T., and O'Keefe, J. (2000). Predictions derived from modelling the hippocampal role in navigation. *Biological Cybernetics*, **83**(3), 301–312.

Burgess, N., Maguire, E., Spiers, H., and O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *NeuroImage*, **14**, 439–453.

Burgess, P., Dumontheil, I., and Gilbert, S. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences*, **11**(7), 290–298.

Burnod, Y., Baraduc, P., Battaglia-Mayer, A., Guigon, E., Koechlin, E., Ferraina, S., Laquaniti, F., and Caminiti, R. (1999). Parieto-frontal coding of reaching: An integrated framework. *Experimental Brain Research*, **129**, 325–346.

Buschman, T. and Miller, E. (2007). Top-down and bottom-up attention in the prefrontal and posterior parietal cortices. *Science*, **315**, 1860–1862.

Butt, M. (1995). *The Structure of Complex Predicates in Urdu*. Center for the Study of Language and Information, Stanford, CA.

Butterworth, B. and Warrington, E. (1995). Two routes to repetition: Evidence from a case of 'deep dysphasia'. *Neurocase*, **1**, 55–66.

Butterworth, G. (2006). Joint visual attention in infancy. In G. Bremner and A. Fogel, editors, *Blackwell Handbook of Infant Development*, pages 213–240. Wiley-Blackwell, Oxford. 2nd edition.

Butterworth, G. and Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms for perspective taking in infancy. *British Journal of Developmental Psychology*, **9**, 55–72.

Buzsàki, G. and Chrobak, J. (1995). Temporal structure in spatially organized neuronal ensembles: A role for interneuronal networks. *Current Opinion in Neurobiology*, **5**(4), 504–510.

Bybee, J. and McClelland, J. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguistic Review*, **22**(2–4), 381–410.

Call, J. (2004). Inferences about the location of food in the great apes (pan paniscus, pan troglodytes, gorilla gorilla and pongo pygmaeus). *Journal of Comparative Psychology*, **118**, 232–241.

Call, J., Hare, B., Carpenter, M., and Tomasello, M. (2004). 'Unwilling' versus 'unable': Chimpanzees' understanding of human intentional action. *Developmental Science*, **7**(4), 488–498.

Calvin, W. and Bickerton, D. (2000). *Lingua ex Machina: Reconciling Darwin and Chomsky with the human brain*. MIT Press, Cambridge, MA.

Caminiti, R., Johnson, P., Galli, C., Ferraina, S., and Burnod, Y. (1991). Making arm movements within different parts of space: The premotor and motor cortical representation of a coordinate system for reaching to visual targets. *Journal of Neuroscience*, **11**(5), 1182–1197.

Caminiti, R., Genovesio, A., Marconi, B., Battaglia-Mayer, A., Onorati, P., Ferraina, S., Mitsuda, T., Giannetti, S., Squatrito, S., Maioli, M., and Molinari, M. (1999). Early coding of reaching: Frontal and parietal association connections of parieto-occipital cortex. *European Journal of Neuroscience*, **11**(9), 3339–3345.

Caplan, C. and Waters, G. (1992). Issues arising regarding the nature and consequences of reproduction conduction aphasia. In S. Kohn, editor, *Conduction Aphasia*, pages 117–150. Laurence Erlbaum Associates, Mahwah, NJ.

Caplan, D., Alpert, N., and Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, **10**(4), 541–552.

Caramazza, A., Capasso, R., Capitani, E., and Miceli, G. (2005). Patterns of comprehension performance in agrammatic Broca's aphasia: A test of the trace deletion hypothesis. *Brain and Language*, **94**(1), 43–53.

Carey, D., Perrett, D., and Oram, M. (1997). Recognising, understanding and reproducing action. In M. Jeannerod, editor, *Handbook of Neuropsychology. Volume 11: Action and Cognition*, pages 111–129. Elsevier, Amsterdam.

Carlson, E. and Triesch, J. (2003). A computational model of the emergence of gaze following. In H. Bowman and C. Labiouse, editors, *Proceedings of the 8th Neural Computation Workshop (NCPW8)*, Progress in Neural Processing. World Scientific.

Carnie, A. (2007). *Syntax: A Generative Introduction*. Blackwell, Oxford.

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, **63**(4), 1–174.

Carrington, S. and Bailey, A. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, **30**, 2313–2335.

Castiello, U. (2003). Understanding other people's actions: Intention and attention. *Journal of Experimental Psychology: Human Perception and Performance*, **29**(2), 416–430.

Castiello, U. and Jeannerod, M. (1991). Measuring time to awareness. *Neuroreport*, **2**(12), 787–800.

Catani, M., Jones, D., and ffytche, D. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, **57**, 8–16.

Caza, G. and Knott, A. (2012). Pragmatic bootstrapping: A neural network model of vocabulary acquisition. *Language Learning and Development*, **8**, 1–23.

Chafee, M., Averbeck, B., and Crowe, D. (2007). Representing spatial relationships in posterior parietal cortex: Single neurons code object-referenced position. *Cerebral Cortex*, **17**(12), 2914–2932.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, **26**, 609–651.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, N. (1980). *Rules and Representations*. Columbia University Press, New York.

Chomsky, N. (1981). *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press, Cambridge, MA.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, **36**(1), 1–22.

Chong, T., Cunnington, R., Williams, M., Kanwisher, N., and Mattingley, J. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology*, **18**, 1576–1580.

Christiansen, M. and Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, **23**, 157–205.

Christiansen, M. and Kirby, S. (2003). Language evolution: The hardest problem in science? In M. Christiansen and S. Kirby, editors, *Language Evolution*. Oxford University Press, Oxford.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective. Oxford:*. Oxford University Press, Oxford, UK.

Cisek, P. (2005). Neural representations of motor plans, desired trajectories, and controlled objects. *Cognitive Processes*, **6**, 15–24.

Cisek, P. and Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, **45**, 801–814.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge, MA.

Clark, H. (1996). *Using Language*. Cambridge University Press, Cambridge, UK.

Cohen, H. and Lefebvre, C. (2005). *Handbook of categorization in cognitive science*. Elsevier, Amsterdam.

Cohen, P. and Levesque, H. (1990a). Rational interaction as the basis for communication. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, pages 221–256. MIT Press, Cambridge, MA.

Cohen, P. and Levesque, H. (1990b). Performatives in a rationally based speech act theory. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, Pittsburgh, PA.

Colby, C. and Goldberg, M. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, **22**, 319–349.

Colby, C., Duhamel, J., and Goldberg, M. (1993). Ventral intraparietal area of the macaque: Anatomic location and visual response properties. *Journal of Neurophysiology*, **69**, 902–914.

Collins, M. (1996). A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Meeting of the ACL*, Santa Cruz.

Conway, C. and Christiansen, M. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, **5**(12), 539–546.

Cooke, D., Charlotte, S., Taylor, T., and Graziano, M. (2003). Complex movements evoked by microstimulation of the ventral intraparietal area. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(10), 6163–6168.

Cooke, S. and Bliss, T. (2006). Plasticity in the human central nervous system. *Brain*, **129**, 1659–1673.

Cooper, R. (1983). *Quantification and syntactic theory*. Reidel, Dordrecht.

Cooper, R. and Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, **17**(4), 297–338.

Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, CA.

Corballis, M. (2002). *From Hand to Mouth: The Origins of Language*. Princeton University Press, Princeton.

Courtney, S., Petit, L., Maisog, J., Ungerleider, L., and Haxby, J. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, **279**(5355), 134–1351.

Craik, F. and Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, **11**, 671–684.

Culham, J., Brandt, S., Cavanagh, P., Kanwisher, N., Dale, A., and Tootell, R. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, **80**, 2657–2670.

Damasio, A., editor (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace, New York.

Damasio, A. and Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In C. Koch and J. Davis, editors, *Large-scale Neuronal Theories of the Brain*. MIT Press, Cambridge, MA.

Damasio, A. and Damsio, H. (1980). The anatomical basis of conduction aphasia. *Brain*, **103**, 337–350.

Damasio, A. and Damsio, H. (1992). Brain and language. *Scientific American*, **267**(3), 88–95.

Damasio, H., Grabowski, T., Tranel, D., Hichwa, R., and Damasio, A. (1996). A neural basis for lexical retrieval. *Nature*, **380**, 499–505.

Davidson, T., Kloosterman, F., and Wilson, M. (2009). Hippocampal replay of extended experience. *Neuron*, **63**, 497–507.

DeGraff, M., editor (2001). *Language Creation and Language Change: Creolization, Diachrony, and Development*. MIT Press, Cambridge, MA.

Dell, G. and Reich, P. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, **20**, 611–629.

Demb, J., Desmond, J., Wagner, A., Vaidya, C., Glover, G., and Gabriel, J. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, **15**(9), 5870–5878.

Demiris, J. and Hayes, G. (2002). Imitation as a dual-route process featuring predictive and learning components: A biologically plausible computational model. In C. Nehaniv and K. Dautenhahn, editors, *Imitation in animals and artifacts*. MIT Press.

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193–222.

Desmurget, M., Epstein, C., Turner, R., Prablanc, C., Alexander, G., and Grafton, S. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nature Neuroscience*, **2**, 563–567.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, **91**, 176–180.

Diana, R., Yonelinas, A., and Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, **11**(9), 379–386.

Diba, K. and Buzsàki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, **10**(10), 1241–1242.

Diesing, M. (1992). *Indefinites*. MIT Press, Cambridge, MA.

Dimitrijevi, L. and Bjelakovi, B. (2004). Development of cardinal motor skills in the first year of life. *Acta Facultatis Medicae Naissensis*, **21**(4), 253–257.

Dixon, R. (1994). *Ergativity*. Cambridge University Press, Cambridge.

Dobbins, I., Foley, H., Schacter, D., and Wagner, A. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron*, **35**(5), 989–996.

Doeller, C. and Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences of the USA*, **105**(15), 5909–5914.

Doeller, C., King, J., and Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences of the USA*, **105**(15), 5915–5920.

Dominey, P. (1997). An anatomically structured sensory-motor sequence learning system displays some general linguistic capacities. *Brain and Language*, **59**, 50–75.

Dominey, P. and Inui, T. (2009). Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling. *Cortex*, **45**(8), 1012–1018.

Dominey, P., Arbib, M., and Joseph, J.-P. (1995). A model of corticostriatal plasticity for learning associations and sequences. *Journal of cognitive neuroscience*, **7**(3), 311–336.

Dominey, P., Hoen, M., Blanc, J. M., and Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: Evidence from simulation, aphasia and ERP studies. *Brain and Language*, **86**, 207–225.

Dominey, P., Hoen, M., and Inui, T. (2006). A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience*, **18**(12), 2088–2107.

Donoghue, J., Leibovic, S., and Sanes, J. (1992). Organization of the forelimb area in squirrel monkey motor cortex: Representation of digit, wrist and elbow muscles. *Experimental Brain Research*, **89**, 1–19.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, **67**(3), 547–619.

Driver, J. (1999). Egocentric and object-based visual neglect. In N. Burgess, K. Jeffery, and J. O'Keefe, editors, *Spatial functions of the hippocampal formation and the parietal cortex*, pages 67–89. Oxford University Press.

Driver, J. and Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, **2**(7), 254–262.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, **6**(5), 509–540.

Dronkers, N., Wilkins, D., Van Valin, R., Redfern, B., and Jaeger, J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, **92**(1–2), 145–177.

Duncan, J. and Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433–458.

Dux, P., Ivanoff, J., Asplund, C., and Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron*, **52**(6), 1109–1120.

Edelman, S. and Christiansen, M. (2003). How seriously should we take Minimalist syntax? *Trends in Cognitive Sciences*, **7**(2), 60–61.

Edelstein, E. (2012). *The Syntax of Adverb Distribution*. Ph.D. thesis, Department of Linguistics and English Language, University of Edinburgh.

Ego-Stengel, V. and Wilson, M. (2010). Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, **20**(1), 1–10.

Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron*, **44**, 109–120.

Eichenbaum, H., Yonelinas, A., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, **30**, 123–152.

Ekstrom, A., Kahana, M., Caplan, J., Fields, T., Isham, E., Newman, E., and Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, **184**, 184–187.

Eldridge, L., Knowlton, B., Furmanski, C., Bookheimer, S., and Engel, S. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, **3**(11), 1149–1152.

Eldridge, L., Engel, S., Zeineh, M., Bookheimer, S., and Knowlton, B. (2005). A dissociation of encoding and retrieval processes in the human hippocampus. *Journal of Neuroscience*, **25**(13), 3280–3286.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, **14**, 179–211.

Elman, J. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, **7**, 195–225.

Elman, J. (1995). Language as a dynamical system. In R. Port and T. van Gelder, editors, *Mind as motion: Explorations in the dynamics of cognition*, pages 195–225. MIT Press.

Emonds, J. (1978). The verbal complex V'–V in French. *Linguistic Inquiry*, **9**, 151–175.

Epstein, R. and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, **392**, 598–601.

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, **23**, 115–125.

Epstein, R., Graham, K., and Downing, P. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, **37**, 865–876.

Epstein, R., Higgins, S., and Thompson-Schill, S. (2005). Learning places from views: Variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience*, **17**(1), 73–83.

Epstein, R., Higgins, S., Jablonski, K., and Feiler, A. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology*, **97**, 3670–3683.

Ernst, T. (2002). *The Syntax of Adjuncts*. Cambridge University Press, Cambridge, UK.

Eskandar, E. and Asaad, J. (1999). Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nature Neuroscience*, **2**(1), 88–93.

Eskandar, E. and Asaad, J. (2002). Distinct nature of directional signals among parietal cortical areas during visual guidance. *Journal of neurophysiology*, **88**, 1777–1790.

Euston, D., Tatsuno, M., and McNaughton, B. (2008). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, **318**, 1147–1150.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation—A magnetic stimulation study. *Journal of Neurophysiology*, **73**(6), 2608–2611.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, **15**, 399–402.

Fadiga, L., Craighero, L., and Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current opinion in neurobiology*, **15**(2), 213–218.

Fagg, A. and Arbib, M. (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks*, **11**(7/8), 1277–1303.

Farrer, C. and Frith, C. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage*, **15**, 596–603.

Farrer, C., Franck, N., Georgieff, N., Frith, C., Decety, J., and Jeannerod, M. (2003). Modulating the experience of agency: A PET study. *NeuroImage*, **18**(2), 324–333.

Feigenson, L., Carey, S., and Hauser, M. (2002). The representations underlying infants' choice of more: Object-files versus analog magnitudes. *Psychological Science*, **13**, 150–156.

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, **8**(7), 307–314.

Feldman, J. and Narayanan, S. (2004). Embodiment in a neural theory of language. *Brain and Language*, **89**(2), 385–392.

Fell, J., Klaver, P., Elfadil, H., Schaller, C., Elger, C., and Fernandez, G. (2003). Rhinal-hippocampal theta coherence during declarative memory formation: interaction with gamma synchronization? *European Journal of Neuroscience*, **17**(5), 1082–1088.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., Pethick, S., Tomasello, M., Mervis, C., and Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59**, 1–185.

Ferbinteanu, J. and Shapiro, M. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron*, **40**, 1227–1239.

Fernando, T. and Kamp, H. (1996). Expecting many. In *Proceedings of Semantics and Linguistic Theory VI*, pages 53–68, Rutgers/Cornell.

Ferraina, S., Paré, M., and Wurtz, R. (2000). Disparity sensitivity of frontal eye field neurons. *Journal of neurophysiology*, **83**, 625–629.

Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *Linguistic Review*, **22**(2–4), 365–380.

Fiebach, C., Schlesewsky, M., and Friederici, A. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, **30**(3), 321–338.

Fillmore, C., Kay, P., and O'Connor, M. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, **64**, 501–38.

Fink, G., Marshall, J., Halligan, P., Frith, C., Driver, J., Frackowiack, R., and Dolan, R. (1999). The neural consequences of conflict between intention and the senses. *Brain*, **122**, 497–512.

Flanagan, J. and Johansson, R. (2003). Action plans used in action observation. *Nature*, **424**, 769–771.

Flanders, M., Tillery, S., and Soechting, J. (1992). Early stages in a sensorimotor transformation. *Behavioral and Brain Sciences*, **15**, 309–320.

Fleck, M., Daselaar, S., Dobbins, I., and Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cerebral Cortex*, **16**, 1623–1630.

Fletcher, P. and Henson, R. (2001). Frontal lobes and human memory—Insights from functional neuroimaging. *Brain*, **124**, 849–881.

Fodor, J. (1983). *The Modularity of Mind*. MIT/Bradford Press, Cambridge, MA.

Fodor, J. and Pylyshyn, Z. (1988). Cognition and cognitive architecture: A critical analysis. *Cognition*, **28**, 3–71.

Fogassi, L., Gallese, V., Fadiga, G., Luppino, G., Matelli, M., and Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of Neurophysiology*, **76**, 141–157.

Fogassi, L., Raos, V., Franchi, G., Gallese, V., Luppino, G., and Matelli, M. (1999). Visual responses in the dorsal premotor area F2 of the macaque monkey. *Experimental Brain Research*, **128**, 194–199.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: From action organisation to intention understanding. *Science*, **308**, 662–667.

Fortin, N., Agster, K., and Eichenbaum, H. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, **5**(5), 458–462.

Foss, D. and Cairns, H. (1970). Some effects of memory limitation upon sentence comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, **9**, 541–547.

Foster, D. and Wilson, M. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, **440**, 680–683.

Foster, D., Morris, R., and Dayan, P. (2000). A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus*, **10**, 1–16.

Frank, L., Brown, E., and Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, **27**(1), 169–178.

Franzius, M., Sprekeler, H., and Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Computational Biology*, **3**(8), 1605–1622.

Freedman, D., Riesenhuber, M., Poggio, T., and Miller, E. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorisation. *Journal of Neuroscience*, **15**, 5235–5246.

Frey, S. and Petrides, M. (2002). Orbitofrontal cortex and memory formation. *Neuron*, **36**, 171–176.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, **6**(2), 78–84.

Friederici, A., Meyer, M., and von Cramon, D. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, **74**, 289–300.

Frischen, A., Loach, D., and Tipper, S. (2009). Seeing the world through another persons eyes: Simulating selective attention via action observation. *Cognition*, **111**, 212–218.

Frith, C. and Frith, U. (2006). The neural basis of mentalizing. *Neuron*, **50**, 531–534.

Fromkin, V. (1973). The non-anomalous nature of anomalous utterances. In V. Fromkin, editor, *Speech errors as linguistic evidence*, pages 215–242. Mouton, The Hague.

Fyhn, M., Molden, S., Witter, M., Moser, E., and Moser, M. (2004). Spatial representation in the entorhinal cortex. *Science*, **305**, 1258–1264.

Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, **2**(12), 493–501.

Gallese, V. and Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, **22**(3/4), 455–479.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, **119**, 593–609.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. In W. Prinz and B. Hommel, editors, *Common Mechanisms in Perception and Action: Attention and Performance 19*, pages 334–355. Oxford University Press, Oxford, UK.

Galletti, C., Kutz, D., Gamberini, M., Breveglieri, R., and Fattori, P. (2003). Role of the medial parieto-occipital cortex in the control of reaching and grasping movements. *Experimental Brain Research*, **153**(2), 158–170.

Gangitano, M., Mottaghy, F., and Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*, **12**(7), 1489–1492.

Gärdenfors, P. (2004). *Conceptual Spaces*. MIT Press, Cambridge, MA.

Gathercole, S. and Baddeley, A. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, **81**, 439–454.

Gentner, D. and Markman, A. (1997). Structure mapping in analogy and similarity. *American Psychologist*, **52**, 45–56.

Georgieff, N. and Jeannerod, M. (1998). Beyond consciousness of external reality: A "who" system for consciousness of action and self-consciousness. *Consciousness and cognition*, **7**, 465–477.

Georgopoulos, A., Kalaska, J., Caminiti, R., and Massey, J. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, **2**(11), 1527–1537.

Gershberg, F. and Shimamura, A. (1995). Impaired use of organizational strategies in free-recall following frontal lobe damage. *Neuropsychologia*, **33**(10), 1305–1333.

Gershkoff-Stowe, L. and Goldin-Meadow, S. (2002). Is there a natural order for expressing semantic relations? *Cognitive Psychology*, **45**, 375–412.

Geuder, W. (2002). *Oriented Adverbs: Issues in the Lexical Semantics of Event Adverbs*. Ph.D. thesis, Faculty of Philology, University of Tübingen.

Geyer, S., Matelli, M., Luppino, G., and Zilles, K. (2000). Functional neuroanatomy of the primate isocortical motor system. *Anatomy and Embryology*, **202**(6), 443–474.

Gibson, J., editor (1950). *The perception of the visual world*. Houghton Mifflin, Boston.

Giese, M. (2000). Neural model for the recognition of biological motion. In G. Baratoff and H. Neumann, editors, *Dynamische Perzeption*, pages 105–110. Infix Verlag, Berlin.

Givón, T. (2002). The visual information-processing system as an evolutionary precursor of human language. In T. Givón and B. Malle, editors, *The evolution of Language out of Prelanguage*. John Benjamins, Amsterdam.

Gleitman, L., January, D., Nappa, R., and Trueswell, J. (2007). On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language*, **57**, 544–569.

Glenberg, A. and Kaschak, P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, **9**(3), 558–565.

Gluck, M. and Myers, C. (2001). *Gateway to Memory: An Introduction to Neural Network Models of the Hippocampus and Learning*. MIT Press, Cambridge, MA.

Gnadt, J. and Mays, L. (1995). Neurons in monkey parietal area LIP are tuned for eye-movement parameters in 3-dimensional space. *Journal of neurophysiology*, **73**(1), 280–297.

Goldberg, A. (1995). *Constructions. A Construction Grammar approach to Argument Structure*. University of Chicago Press, Chicago.

Goldin-Meadow, S., So, W., Özürek, A., and Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the USA*, **105**(27), 9163–9168.

Goldstein, L. and Fowler, C. (2003). Articulatory phonology: A phonology for public language use. In A. Meyer and N. Schiller, editors, *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, pages 159–207. Mouton, New York.

Gomi, H., Shidara, M., Takemura, A., Inoue, Y., Kawano, K., and Kawato, M. (1998). Temporal firing patterns of purkinje cells in the cerebellar ventral paraflocculus during ocular following responses in monkeys. I. Simple spikes. *Journal of Neurophysiology*, **80**, 818–831.

Goodwin, A. and Wheat, H. (2004). Sensory signals in neural populations underlying tactile perception and manipulation. *Annual Review of Neuroscience*, **27**, 53–77.

Goschke, T., Friederici, A., Kotz, S., and van Kampen, A. (2001). Procedural learning in Broca's aphasia: Dissociation between the implicit acquisition of spatio-motor and phoneme sequences. *Journal of Cognitive Neuroscience*, **13**(3), 370–388.

Gottlieb, J., Kusunoki, M., and Goldberg, M. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, **391**, 481–484.

Gough, P., Nobre, A., and Devlin, J. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *Journal of Neuroscience*, **25**(35), 8010–8016.

Grafton, S. and Hamilton, A. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, **26**, 590–616.

Graves, W., Grabowski, T., Mehta, S., and Gupta, P. (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of Cognitive Neuroscience*, **20**(9), 1698–1710.

Graybiel, A. and Saka, E. (2004). The basal ganglia and the control of action. In M. Gazzaniga, editor, *The Cognitive Neurosciences III*, pages 495–510. MIT Press, Cambridge, MA.

Greenfield, P. (1991). Language, tools and the brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, **14**, 531–535.

Greenlee, M., Magnussen, S., and Reinvang, I. (2000). Brain regions involved in spatial frequency discrimination: Evidence from fMRI. *Experimental Brain Research*, **132**, 399–403.

Grèzes, J. and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation and verb generation of actions: A meta-analysis. *Human Brain Mapping*, **12**(1), 1–19.

Griffin, Z. (2004). Why look? reasons for eye movements related to language production. In J. Henderson and F. Ferreira, editors, *The Interface of Language, Vision and Action: Eye Movements and the Visual World*, pages 213–248. Psychology Press, New York.

Griffin, Z. and Bock, K. (2000). What the eyes say about speaking. *Psychological science*, **11**, 274–279.

Grodzinsky, J. and Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences*, **12**(12), 474–480.

Grosbras, M. and Paus, T. (2003). Transcranial magnetic stimulation of the human frontal eye field facilitates visual awareness. *European Journal of Neuroscience*, **18**(11), 3121–3126.

Grossberg, S. (1978). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, **17**, 199–219.

Grossman, E. and Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, **35**, 1167–1175.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., and Blake, R. (2000). Brain areas involved in the perception of biological motion. *Journal of Cognitive Neuroscience*, **12**(5), 711–719.

Grosu, A. (1988). On the distribution of genitive phrases in Romanian. *Linguistics*, **26**, 931–949.

Haegeman, L. (1991). *Introduction to Government and Binding Theory*. Blackwell, Oxford.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, **9**(9), 416–423.

Hamilton, A. and Grafton, S. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, **26**(4), 1133–1137.

Hamker, F. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research*, **44**, 501–521.

Hare, B. and Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, **68**, 571–581.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., and Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(25), 15061–15065.

Harley, H. (2003). Possession and the double object construction. In P. Pica and J. Rooryck, editors, *Linguistic Variation Yearbook Vol. 2 (2002)*, pages 31–70. John Benjamins.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, **42**, 335–346.

Harris, C. and Wolpert, D. (1998). Signal-dependent noise determines motor planning. *Nature*, **394**(6695), 780–784.

Harris, Z. (1954). Distributional structure. *Word*, **10**(23), 146–162.

Hartley, T. and Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, **35**, 1–31.

Hartley, T., Burgess, N., Lever, C., Cacucci, F., and O'Keefe, J. (2000). Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus*, **10**(4), 369–379.

Hartley, T., Maguire, E., Spiers, H., and Burgess, N. (2003). The well-worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans. *Neuron*, **37**, 877–888.

Hasegawa, R., Matsumoto, M., and Mikami, A. (2000). Search target selection in monkey prefrontal cortex. *Journal of Neurophysiology*, **84**, 1692–1696.

Hashimoto, R. and Sakai, K. (2002). Specialization in the left prefrontal cortex for sentence comprehension. *Neuron*, **35**, 589–597.

Hauf, P., Elsner, B., and Aschersleben, G. (2004). The role of action effects in infants' action control. *Psychological Research*, **68**, 115–125.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, **41**, 301–307.

Haxby, J., Petit, L., Ungerleider, L., and Courtney, S. (2000). Distinguishing the functional roles of multiple regions in distributed neural systems for visual working memory. *NeuroImage*, **11**, 145–156.

Hebb, D. (1949). *The organization of behavior*. J Wiley and Sons, New York.

Hebb, D. (1961). Distinctive features of learning in the higher animal. In J. Delafresnaye, editor, *Brain Mechanisms and Learning*, pages 37–46. Oxford University Press.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Ph.D. thesis, University of Massachusetts. Distributed by Graduate Linguistic Student Association.

Henderson, J. and Ferreira, F. (2004). Scene perception for psycholinguists. In J. Henderson and F. Ferreira, editors, *The Interface of Language, Vision and Action: Eye Movements and the Visual World*, pages 1–58. Psychology Press, New York.

Henriques, D., Flanders, M., and Soechting, J. (2004). Haptic synthesis of shapes and sequences. *Journal of Neurophysiology*, **91**, 1808–1821.

Henson, R., Rugg, M., Shallice, T., Josephs, O., and Dolan, R. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, **19**(10), 3962–3972.

Henson, R., Burgess, N., and Frith, C. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia*, **38**(4), 426–440.

Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, **92**, 67–99.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, **8**(5), 393–402.

Hickok, G., Zurif, E., and Canseco-Gonzalez, E. (1993). Structural description of agrammatic comprehension. *Brain and Language*, **45**, 371–395.

Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A., Naeve-Velguth, S., Struppf, J., Strick, P., and Ugurbil, K. (2000). A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: Implications for the explanation of conduction aphasia. *Neuroscience Letters*, **287**, 156–160.

Hirsh-Pasek, K. and Golinkoff, R. (1996). *The Origins of Grammar: Evidence from Early Language Comprehension*. MIT Press, Cambridge, MA.

Hirtle, S. and Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and Cognition*, **13**(3), 208–217.

Hockenmaier, J. and Steedman, M. (2002). Generative models for statistical parsing with combinatory grammars. In *Proceedings of the 40th meeting of the Association for Computational Linguistics*, pages 335–342, Philadelphia.

Hoen, M., Pachot-Clouard, M., Segebarth, C., and Dominey, P. (2006). When Broca experiences the Janus syndrome. An ER-fMRI study comparing sentence comprehension and cognitive sequence processing. *Cortex*, **42**, 605–623.

Hoffman, K. and McNaughton, B. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science*, **297**, 2070–2073.

Hok, V., Save, E., Lenck-Santini, P., and Poucet, B. (2005). Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proceedings of the National Academy of Sciences*, **102**(12), 4602–4607.

Hok, V., Lenck-Santini, P., Roux, S., Save, E., Muller, R., and Poucet, B. (2007). Goal-related activity in hippocampal place cells. *Journal of Neuroscience*, **27**(3), 472–482.

Holdstock, J., Mayes, A., Gong, Q., Roberts, N., and Kapur, N. (2005). Item recognition is less impaired than recall and associative recognition in a patient with selective hippocampal damage. *Hippocampus*, **5**, 203–215.

Hollup, S., Molden, S., Donnett, J., Moser, M., and Moser, E. (2001). Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, **21**, 1635–1644.

Holmes, G. (1939). The cerebellum of man. *Brain*, **62**, 1–30.

Hölscher, C. (2003). Time, space and hippocampal functions. *Reviews in the Neurosciences*, **14**(3), 253–284.

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action learning. *Behavioral and Brain Sciences*, **24**, 849–878.

Hooker, C., Paller, K., Gitelman, D., Parrish, T., Mesulam, M., and Reber, P. (2003). Brain networks for analyzing eye gaze. *Cognitive Brain Research*, **17**(2), 406–418.

Hornstein, N. (2009). *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge University Press, Cambridge, UK.

Houghton, G. and Hartley, T. (1995). Parallel models of serial behaviour: Lashley revisited. *Psyche*, **2**(25).

Howard, M. and Kahana, M. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, **46**, 269–299.

Hughes, H., Nozawa, G., and Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, **8**(3), 197–230.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G., Martin, S., and Stuart, G. (1991). Word-frequency effects on short-term memory tasks: Evidence for a redintigration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **23**(5), 1217–1232.

Hummel, J. and Holyoak, K. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, **110**(2), 220–264.

Humphries, C., Willard, K., Buchsbaum, B., and Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *Neuroreport*, **12**(8), 1749–1752.

Hurford, J. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, **26**(3), 261–283.

Huyck, C. (2007). Creating hierarchical categories using cell assemblies. *Neural Computation*, **19**(1), 1–24.

Iacoboni, M., Woods, R., Brass, M., Bekkering, H., Mazziotta, J., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, **286**, 2526–2528.

Iacoboni, M., Koski, L., Brass, M., Bekkering, H., Woods, R., Dubeau, M., Mazziotta, J., and Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(24), 13995–13999.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, **3**(3), e79.

Iatridou, S. (1990). About Agr(P). *Linguistic Inquiry*, **21**(4), 421–459.

Iberall, T. and Arbib, M. (1990). Schemas for the control of hand movements: An essay on cortical localisation. In M. Goodale, editor, *Vision and Action: The Control of Grasping*, pages 163–180. Ablex, Norwood.

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Putz, B., Yoshioka, T., and Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, **403**(6766), 192–195.

Indefrey, P., Brown, C., Hellwig, F., Amunts, K., Herzog, H., Seitz, R., and Hagoort, P. (2001). A neural correlate of syntactic encoding during speech production. *Proceedings of the National Academy of Sciences*, **98**(10), 5933–5936.

Ito, T., Tiede, M., and Ostry, D. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, **106**(4), 1245–1248.

Itti, L. and Arbib, M. (2006). Attention and the minimal subscene. In M. Arbib, editor, *Action to Language via the Mirror Neuron System*, pages 289–346. Cambridge University Press, Cambridge, UK.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, **40**(10–12), 1489–1506.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews—Neuroscience*, **2**, 1–11.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, MA.

Jackendoff, R. (1977). *X-Bar Syntax: A Study of Phrase Structure.*. Linguistic Inquiry Monograph 2. MIT Press, Cambridge, MA.

Jackendoff, R. (1999). The architecture of the linguistic-spatial interface. In P. Bloom, M. Peterson, L. Nadel, and M. Garrett, editors, *Language and Space*, pages 1–30. MIT Press, Cambridge, MA.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford.

Jackendoff, R. (2007). Linguistics in cognitive science: The state of the art. *The Linguistic Review*, **24**, 347–401.

Jacquemot, C. and Scott, S. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, **10**(11), 480–486.

Jankelowitz, S. and Colebatch, J. (2002). Movement-related potentials associated with self-paced, cued and imagined arm movements. *Experimental Brain Research*, **147**(1), 98–107.

Jeannerod, M. (1996). Reaching and grasping. Parallel specification of visuomotor channels. In W. Prinz and B. Bridgeman, editors, *Handbook of Perception and Action Volume II: Motor Skills*, pages 405–460. Academic Press.

Jeannerod, M. (1999). To act or not to act: Perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology: A*, **52**(1), 1–29.

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, **14**, S103–S109.

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain Research*, **142**, 1–15.

Jeffery, K. (2007). Integration of the sensory inputs to place cells: what, where, why, and how? *Hippocampus*, **17**, 775–785.

Jeffery, K., Gilbert, A., Burton, S., and Strudwick, A. (2003). Preserved performance in a hippocampal-dependent spatial task despite complete place cell remapping. *Hippocampus*, **13**, 175–189.

Jellema, T., Baker, C., Wicker, B., and Perrett, D. (2000). Neural representation for the perception of the intentionality of actions. *Brain and Cognition*, **44**, 280–302.

Jellema, T., Maassen, G., and Perrett, D. (2004). Single cell integration of animate form, motion and location in the superior temporal cortex of the macaque monkey. *Cerebral Cortex*, **14**(7), 781–790.

Jenkins, H., Jahanshahi, M., Jueptner, M., Passingham, R., and Brooks, D. (2000). Self-initiated versus externally triggered movements. II. the effect of movement predictability on regional cerebral blood flow. *Brain*, **123**(6), 1216–1228.

Jensen, O. and Lisman, J. (1996). Hippocampal CA3 region predicts memory sequences: Accounting for the phase precession of place cells. *Learning and Memory*, **3**(2–3), 279–287.

Jensen, O. and Lisman, J. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends in Neurosciences*, **28**(2), 67–72.

Ji, D. and Wilson, M. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, **10**(1), 100–107.

Ji, D. and Wilson, M. (2008). Firing rate dynamics in the hippocampus induced by trajectory learning. *Journal of Neuroscience*, **21**(18), 4679–4689.

Jiang, H. (2005). Confidence measures for speech recognition: A survey. *Speech Communication*, **45**, 455–470.

Johansson, G. (1973). Visual perception of biological motion, and a model for its analysis. *Visual Perception and Psychophysics*, **14**, 201–211.

Johansson, M. and Mecklinger, A. (2003). The late posterior negativity in ERP studies of episodic memory: Action monitoring and retrieval of attribute conjunctions. *Biological Psychology*, **64**(1–2), 91–117.

Johansson, R., Westling, G., Backstrom, A., and Flanagan, J. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, **21**(17), 6917–6932.

Johnson, A. and Redish, D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, **27**(45), 12176–12189.

Johnson-Laird, P. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge.

Jordan, M. and Wolpert, D. (2000). Computational motor control. In M. Gazzaniga, editor, *The New Cognitive Neurosciences*, pages 71–118. MIT Press.

Jung, M., Wiener, S., and McNaughton, B. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *Journal of Neuroscience*, **14**, 7347–7356.

Justus, T. (2004). The cerebellum and English grammatical morphology: Evidence from production, comprehension and grammatical morphology judgements. *Journal of Cognitive Neuroscience*, **16**(7), 1115–1130.

Kahneman, D., Treisman, A., and Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175–219.

Kakei, S., Hoffman, D., and Strick, P. (2001). Direction of action is represented in the ventral premotor cortex. *Nature Neuroscience*, **4**(10), 1020–1025.

Kalaska, J., Sergio, L., and Cisek, P. (1998). Cortical control of whole-arm motor tasks. In M. Glickstein, editor, *Sensory Guidance of Movement. Novartis Foundation Symposium No. 218*, pages 176–201. Wiley, Chichester, UK.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof, editors, *Formal Methods in the Study of Language*, pages 277–322. Mathematical Center Tract 135, Amsterdam.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht.

Kawato, M. and Gomi, H. (1992). The cerebellum and VOR/OKR learning models. *Trends in Neurosciences*, **15**, 445–453.

Kawato, M., Furawaka, K., and Suzuki, R. (1987). A hierarchical neural network model for the control and learning of voluntary movements. *Biological Cybernetics*, **56**, 1–17.

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S., and Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. In C. Prablanc, D. Pélisson, and Y. Rosetti, editors, *Progress in Brain Research*, volume 142, pages 171–188. Elsevier Science.

Kay, J. and Ellis, A. (1987). A cognitive neuropsychological case study of anomia: Implications for psychological models of word retrieval. *Brain*, **110**, 613–629.

Kayne, R. (1994). *The Antisymmetry of Syntax*. MIT Press, Cambridge, MA.

Kazanovich, Y. and Borisyuk, R. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation*, **18**, 1413–1440.

Kegl, J., Senghas, A., and Coppola, M. (2001). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. DeGraff, editor, *Language Creation and Language Change: Creolization, Diachrony, and Development*. MIT Press, Cambridge, MA.

Kellogg, R. (2003). *Cognitive Psychology (2nd edition)*. Sage Publication, Thousand Oaks, CA.

Kelso, K., Southard, D., and Goodman, D. (1979). Nature of human inter-limb coordination. *Science*, **203**(4384), 1029–1031.

Kent, C. and Lamberts, K. (2008). The encoding-retrieval relationship: Retrieval as mental simulation. *Trends in Cognitive Sciences*, **12**(3), 92–98.

Kesner, R., Gilbert, P., and Barua, L. (2002). The role of the hippocampus in memory for the temporal order of a sequence of odors. *Behavioral Neuroscience*, **116**(2), 286–290.

Keysers, C. and Perrett, D. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, **8**(11), 501–507.

Keysers, C., Wicker, B., Gazzola, V., Anton, J., Fogassi, L., and Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, **42**, 335–346.

Kiani, R. and Shadlen, M. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, **324**, 759–764.

Kingstone, A., Friesen, C., and Gazzaniga, M. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science*, **11**(2), 159–166.

Kirby, S. and Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi, editors, *Simulating the Evolution of Language*. Springer-Verlag, London.

Kjelstrup, J., Solstad, T., Brun, V., Fyhn, M., and Hafting, T. (2007). Very large place fields at the ventral pole of the hippocampal CA3 area. *Soc. Neurosci. Abstr.*, **33.93.1**, 305–314.

Klam, F. and Graf, W. (2006). Discrimination between active and passive head movements by macaque ventral and medial intraparietal cortex neurons. *Journal of Physiology*, **574**(2), 367–386.

Knoblich, G. and Flach, R. (2001). Predicting the effects of actions: interactions of perception and action. *Psychological Science*, **12**(6), 467–472.

Knott, A. (2010). A sensorimotor characterisation of syntax, and its implications for models of language evolution. In *Proceedings of the 8th International conference on the Evolution of Language (EVOLANG 8)*, pages 214–221, New Jersey. World Scientific.

Knott, A. (2014). How infants learn word meanings and propositional attitudes: a neural network model. In T.-W. Hung, editor, *Language and Action*. Springer, Berlin/Heidelberg.

Kobayashi, T., Nishijo, H., Fukuda, M., Bures, J., and Ono, T. (1997). Task-dependent representations in rat hippocampal place neurons. *Journal of Neurophysiology*, **78**, 597–613.

Kobayashi, Y., Kawano, K., Takemura, A., Inoue, Y., Kitama, T., Gomi, H., and Kawato, M. (1998). Temporal firing patterns of purkinje cells in the cerebellar ventral paraflocculus during ocular following responses in monkeys. II. Complex spikes. *Journal of Neurophysiology*, **80**, 832–848.

Koch, C. and Ullman, S. (1985). Shifts in underlying visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, **4**(4), 219–227.

Koffka, K. (1935/2001). *Principles of Gestalt psychology*. Routledge, London.

Kolers, P. and Roediger, H. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, **23**, 425–449.

Koopman, H. and Sportiche, D. (1991). The position of subjects. *Lingua*, **85**, 211–258.

Kosslyn, S., Thompson, W., Kim, I., and Alpert, M. (1995). Topographical representations of mental images in primary visual-cortex. *Nature*, **378**, 496–498.

Kosslyn, S., Thompson, W., and Ganis, G. (2006). *The case for mental imagery*. Oxford University Press, New York.

Köteles, K., De Mazière, P., Van Hulle, M., and Orban, G andVogels, R. (2008). Coding of images of materials by macaque inferior temporal cortical neurons. *European Journal of Neuroscience*, **27**, 466–482.

Kourtzi, Z. and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital cortex. *Science*, **293**, 1506–1509.

Kowler, E., Anderson, E., Dosher, B., and Blaser, E. (1995). The role of attention in the programming of saccdes. *Vision Research*, **35**, 1897–1916.

Kravitz, D., Vinson, L., and Baker, C. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences*, **12**(3), 114–122.

Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, **3**(9), 946–953.

Kumaran, D. and Maguire, E. (2006). The dynamics of hippocampal activation during encoding of overlapping sequences. *Neuron*, **49**, 617–629.

Kusunoki, M., Gottlieb, J., and Goldberg, M. (2000). The lateral intraparietal area as a salience map: The representation of abrupt onset, stimulus motion, and task relevance. *Vision Research*, **40**(10–12), 1459–1468.

Lackner, J. (1988). Some proprioceptive influences on the perceptual representation of body shape and orientation. *Brain*, **111**, 281–297.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*. University of Chicago Press, Chicago and London.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press, Chicago and London.

Land, M. and Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London B*, **352**, 1231–1239.

Land, M., Mennie, N., and Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, **28**, 1311–1328.

Landau, B. and Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press, Cambridge, MA.

Landau, B. and Jackendoff, R. (1993). "What" and "where" in spatial language and cognition. *Behavioral and Brain Sciences*, **16**, 217–265.

Landy, M. and Graham, N. (2001). Visual perception of texture. In L. Chalupa and J. Werner, editors, *The visual neurosciences*, pages ??–?? MIT Press.

Langacker, R. (1987). *Foundations of Cognitive Grammar I: Theoretical Prerequisites*. Stanford University Press, Stanford, CA.

Langacker, R. (1991). *Foundations of Cognitive Grammar II: Descriptive Application*. Stanford University Press, Stanford, CA.

Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press, New York.

Larson, R. (1988). On the double object construction. *Linguistic Inquiry*, **19**(3), 335–391.

Le Cun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. (1992). Handwritten digit recognition with a back-propagation network. In P. Lisboa, editor, *Neural networks: current applications*. Chapman and Hall, London.

Lee, A. and Wilson, M. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, **36**, 1183–1194.

Lee-Hand, J. and Knott, A. (2013). A model of causative actions: from motor learning to syntactic structure. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Berlin.

Leff, A., Schofield, T., Crinion, J., Seghier, M., Grogan, A., Green, D., and Price, C. (2009). The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: Evidence from 210 patients with stroke. *Brain*, **132**(12), 3401–3410.

Leslie, A. (1984). Infant perception of a pick up event. *British Journal of Developmental Psychology*, **2**, 19–32.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

Levelt, W., Roelofs, A., and Meyer, A. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, **22**(1), 1–75.

Levinson, S. (1999). Frames of reference and Molyneux's Question: Crosslinguistic evidence. In P. Bloom, M. Peterson, L. Nadel, and M. Garrett, editors, *Language and Space*, pages 109–169. MIT Press, Cambridge, MA.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Levy, W. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, **6**(6), 579–590.

Li, P., Barner, D., Ogura, T., Yang, S.-J., and Carey, S. (2009). Does the conceptual distinction between singular and plural sets depend on language? *Developmental Psychology*, **45**(6), 1644–1653.

Liberman, A. and Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1–36.

Liddle, M. (2010). *Some neuro-computational investigations into the reviewing of object-files*. Ph.D. thesis, Department of Computer Science, University of Otago.

Lidz, J., Gleitman, H., and Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, **87**, 151–178.

Lieberman, P. (2005). The pied piper of Cambridge. *Linguistic Review*, **22**(2–4), 289–301.

Lieven, E., Pine, J., and Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, **24**, 187–219.

Lin, D. and Pantel, P. (2001). Induction of semantic classes from natural language text. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 317–322.

Lindfield, K., Wingfield, A., and Goodglass, H. (1999). The role of prosody in the mental lexicon. *Brain and Language*, **68**, 312–317.

Lisman, J. and Otmakhova, N. (2001). Storage, recall, and novelty detection of sequences by the hippocampus: Elaborating on the SOCRATIC model to account for normal and aberrant effects of dopamine. *Hippocampus*, **11**(5), 551–568.

Livingstone, M. and Hubel, D. (1988). Segregation of form, color, movement and depth: Anatomy, physiology and perception. *Science*, **240**(4583), 740–749.

Logothetis, N. and Sheinberg, D. (1996). Visual object recognition. *Annual Review of Neuroscience*, **19**, 577–621.

Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, **5**(5), 552–563.

Longobardi, G. (1994). Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry*, **25**(4), 609–665.

Lotto, A., Hickok, G., and Holt, L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, **13**(3), 110–114.

Lotze, M., Montoya, P., Erb, M., Hülsmann, E., Flor, H., Klose, U., Birbaumer, N., and Grodd, W. (1999). Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fMRI study. *Journal of Cognitive Neuroscience*, **11**(5), 491–501.

Lu, L., Crosson, B., Nadeau, S., Heilman, K., Gonzalez-Rothi, L., Raymer, A., Gilmore, R., Bauer, R., and Roper, S. (2002). Category-specific naming deficits for objects and actions: Semantic attribute and grammatical role hypotheses. *Neuropsychologia*, **40**, 1608–1621.

Lumsden, J. (1989). On the distribution of determiners in Haitian Creole. *Revue qubcoise de linguistique*, **18**(2), 65–93.

M, H., Carey, S., and Hauser, L. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London (Series B Biological Sciences)*, **267**, 829–833.

Macaluso, E., Frith, C., and Driver, J. (2002). Supramodal effects of covert spatial orienting triggered by visual or tactile events. *Journal of Cognitive Neuroscience*, **14**(3), 389–401.

MacEvoy, S. and Epstein, R. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, **19**, 943–947.

MacNeilage, P. and Davis, B. (2005). Functional organization of speech across the life span: A critique of generative phonology. *Linguistic Review*, **22**(2–4), 161–181.

Maguire, E., Frackowiak, R., and Frith, C. (1997). Recalling routes around London: Activation of the right hippocampus in taxi drivers. *Journal of Neuroscience*, **17**(19), 7103–7110.

Maioli, M., Squatrito, S., Samolsky-Dekel, B., and Sanseverino, E. (1998). Corticocortical connections between frontal periarcuate regions and visual areas of the superior temporal sulcus and the adjoining inferior parietal lobule in the macaque monkey. *Brain Research*, **789**(1), 118–125.

Manns, J. and Eichenbaum, H. (2006). Evolution of declarative memory. *Hippocampus*, **16**, 795–808.

Marconi, B., Genovesio, A., Battaglia-Mayer, A., Ferraina, S., Squatrito, S., Molinari, M., Lacquaniti, F., and Caminiti, R. (2001). Eye-hand coordination during reaching. I. Anatomical relationships between parietal and frontal cortex. *Cerebral Cortex*, **11**(6), 513–527.

Marois, R. and Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, **9**, 296–305.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical transactions of the Royal Society of London B*, **262**(841), 23–81.

Maunsell, J. and Cook, E. (2002). The role of attention in visual processing. *Philosophical Transactions of the Royal Society of London B*, **357**, 1063–1072.

Maunsell, J. and van Essen, D. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. 1. Selectivity for stimulus direction, speed and orientation. *Journal of neurophysiology*, **49**(5), 1127–1147.

Maylor, E. and Hockey, R. (1985). Inhibitory component of externally controlled covert orienting in visual space. *Journal of Experimental Psychology: Human Perception and Performance*, **11**, 777–786.

Mayr, U. and Keele, S. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, **129**(1), 4–26.

Mayr, U., Diedrichsen, J., Ivry, R., and Keele, S. (2006). Dissociating task-set selection from task-set inhibition in the prefrontal cortex. *Journal of Cognitive Neuroscience*, **18**(1), 14–21.

McClelland, J., Rumelhart, D., and the PDP research group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, volume 2*. MIT Press, Cambridge, MA.

McClelland, J., McNaughton, B., and O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**, 419–457.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746–748.

Melinger, A. and Dobel, C. (2005). Lexically-driven syntactic priming. *Cognition*, **98**, B11–B20.

Meyer, A. (2004). The use of eye tracking in studies of sentence generation. In J. Henderson and F. Ferreira, editors, *The Interface of Language, Vision and Action: Eye Movements and the Visual World*, pages 191–211. Psychology Press, New York.

Miall, C. (2003). Connecting mirror neurons and forward models. *Neuroreport*, **14**(17), 2135–2137.

Miller, E. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, **1**, 59–65.

Miller, E. and Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, **24**, 167–202.

Miller, E. and Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, **263**(5146), 520–522.

Miller, E., Li, L., and Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term-memory task. *Journal of Neuroscience*, **13**(4), 1460–1478.

Miller, G. and Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control*, **7**, 292–303.

Milner, R. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, **27**, 272–277.

Milner, R. and Goodale, M. (1995). *The Visual Brain in Action*. Oxford University Press, Oxford.

Milner, R., Johnsrude, I., and Crane, J. (1997). Right medial temporal-lobe contribution to object-location memory. *Philosophical Transactions of the Royal Society of London B*, **352**, 1469–1474.

Milsark, G. (1974). *Existential sentences in English*. Ph.D. thesis, MIT, Cambridge, MA.

Mitchell, J., Macrae, C., and Banaji, M. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, **50**(4), 655–663.

Mitchell, T., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., and Just, M. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, **320**, 1191–1195.

Moens, M. and Steedman, M. (1987). Temporal ontology in natural language. In *Proceedings of the 25th ACL Conference*, pages 1–7, Stanford, CA.

Moens, M. and Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, **14**(2), 15–28.

Molinari, M., Leggio, M., Solida, A., Ciorra, R., Mischiagna, M., Silveri, M., and Petrosini, L. (1997). Cerebellum and procedural learning: Evidence from focal cerebellar lesions. *Brain*, **120**, 1753–62.

Moore, T. and Armstrong, K. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, **421**(6921), 370–373.

Mozer, M. and Sitton, M. (1996). Computational modeling of spatial attention. In H. Pashler, editor, *Attention*. UCL Press, London.

Mukamel, R., Ekstrom, A., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, **20**(8), 750–756.

Müller, R. and Basho, S. (2004). Are nonlinguistic functions in "Broca's area" prerequisites for language acquisition? fMRI findings from an ontogenetic viewpoint. *Brain and Language*, **89**(2), 329–336.

Muller, R. and Kube, J. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, **7**, 1951–1968.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., and Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*, **78**, 2226–2230.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology*, **83**(5), 2580–2601.

Muskens, R., van Benthem, J., and Visser, A. (1997). Dynamics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 587–648. Elsevier, Amsterdam.

Myachykov, A. and Tomlin, R. (2008). Perceptual priming and structural choice in Russian sentence production. *Cognitive Science*, **6**(1), 31–48.

Myachykov, A., Garrod, S., and Scheepers, C. (2009). Attention and syntax in sentence production: A critical review. *Discours*, **4**, 1–17.

Nádasdy, Z and Hirase, H and Csicsvari, J and Buzsáki, G (1999). Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, **19**(21), 9497–9507.

Nadel, L. and Bohbot, B. (2001). Consolidation of memory. *Hippocampus*, **11**, 56–60.

Nader, K. (2003). Memory traces unbound. *Trends in Neurosciences*, **26**(2), 65–72.

Narayanan, S. (1997). *Knowledge-Based Action Representations for Metaphor and Aspect*. Ph.D. thesis, Department of Computer Science, University of California at Berkeley.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, **9**, 353–383.

Needham, A. and Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, **7**(2), 121–148.

Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G. (2005). Observing others: Multiple action representation in the frontal lobe. *Science*, **310**(5746), 332–336.

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Nieder, A. and Miller, E. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(19), 7457–7462.

Nieder, A., Freedman, D., and Miller, E. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, **297**, 1708–1711.

Nobre, A., Gitelman, D., Dias, E., and Mesulam, M. (2000). Covert visual spatial orienting and saccades: Overlapping neural systems. *NeuroImage*, **11**, 210–216.

Norman, D. and Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. Davidson, G. Schwartz, and D. Shapiro, editors, *Consciousness and Self-Regulation*, volume 4. Plenum.

Noton, D. and Stark, L. (1971). Scan paths in saccadic eye movements while viewing and recognising patterns. *Vision Research*, **11**, 929–944.

Novick, J., Trueswell, J., and Thompson-Schill, S. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective and Behavioural Neuroscience*, **5**(3), 263–281.

Nunberg, G., Sag, I., and Wasow, T. (1994). Idioms. *Language*, **70**, 491–538.

Okada, K. and Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, **98**, 112–117.

O'Keefe, J. (1999). The spatial prepositions in English, Vector Grammar and the Cognitive Map theory. In P. Bloom, M. Peterson, L. Nadel, and M. Garrett, editors, *Language and Space*, pages 277–316. MIT Press, Cambridge, MA.

O'Keefe, J. and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurones. *Nature*, **381**, 425–428.

O'Keefe, J. and Nadel, L. (1978). *The hippocampus as a Cognitive Map*. Clarendon Press, Oxford.

op de Beeck, H. and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, **426**, 505–518.

Oram, M. and Perrett, D. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience*, **6**(2), 99–116.

Oram, M. and Perrett, D. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the Macaque monkey. *Journal of Neurophysiology*, **76**(1), 109–129.

O'Reilly, R. and Frank, M. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, **18**, 283–328.

Osgood, C. (1971). Where do sentences come from? In D. Steinberg, editor, *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, pages 497–529. Cambridge University Press.

Osgood, C. and Bock, J. (1977). Salience and sentencing: Some production principles. In S. Rosenberg, editor, *Sentence Production: Developments in Research and Theory*, pages 89–140. Erlbaum, New Jersey.

Oztop, E. and Arbib, M. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, **87**, 116–140.

Oztop, E., Wolpert, D., and Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, **22**, 129–151.

Paccalin, C. and Jeannerod, M. (2000). Changes in breathing during observation of effortful actions. *Brain Research*, **862**, 194–200.

Packard, M. and McGaugh, J. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, **65**, 65–72.

Paller, K. and Wagner, A. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, **6**(2), 93–102.

Pantel, P. and Lin, D. (2002). Discovering word senses from text. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 613–619.

Papagno, C., Della Salla, S., and Basso, A. (1993). Ideomotor apraxia without aphasia and aphasia without apraxia: The anatomical support for a double dissociation. *Journal of Neurology, Neurosurgery, and Psychiatry*, **56**, 286–289.

Pashler, H., editor (1998). *The psychology of Attention*. MIT Press, Cambridge, MA.

Paulesu, E., Frith, C., and Frackowiak, R. (1993). The neural correlates of the verbal component of working memory. *Nature*, **362**, 342–345.

Pearce, E. (1997). Negation and indefinites in Māori. In D. Forget, P. Hirschbühler, F. Martineau, and M. Rivero, editors, *Negation and Polarity*, pages 271–288. John Benjamins, Amsterdam.

Pelphrey, K., Morris, J., Michelich, C., Allison, T., and McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: An fMRI study of eye, mouth and hand movements. *Cerebral Cortex*, **15**, 1866–1876.

Perani, D., Cappa, S., Schnur, T., Tettamanti, M., Collina, S., Rosa, M., and Fazio, F. (1999). The neural correlates of verb and noun processing—A PET study. *Brain*, **122**, 2337–2344.

Perner, J. (1991). *Understanding the Representational Mind*. MIT Press, Cambridge.

Perrett, D., Smith, P., Potter, D., Mistlin, A., Head, A., Milner, A., and Jeeves, M. (1985). Visual cells in the temporal cortex sensitive to face, view and gaze direction. *Philosophical Transactions of the Royal Society of London B*, **223**, 293–317.

Perrett, D., Harries, M., Bevan, R., Thomas, S., Benson, P., Mistlin, A., Chitty, A., Hiatenen, J., and Ortega, J. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology*, **146**, 87–113.

Perrett, D., Mistlin, A., Harries, M., and Chitty, M. (1990). Understanding the visual appearance and consequence of hand actions. In M. Goodale, editor, *Vision and Action: The Control of Grasping*, pages 162–180. Ablex, Norwood.

Perrett, D., Hietanen, J., Oram, M., and Benson, P. (1992). Organisation and function of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London B*, **335**, 23–30.

Perrone, J. (1992). Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America*, **9**(2), 177–194.

Petrides, M. (1991). Functional specialization within the dorsolateral frontal cortex for serial order memory. *Philosophical Transactions of the Royal Society of London B*, **246**, 299–306.

Petrides, M. and Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, **20**, 249–262.

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S., and Battaglia, F. (2010). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, **12**, 919–926.

Phillips, W. and Christie, D. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, **29**, 117–133.

Piattelli-Palmarini, M., editor (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Harvard University Press, Cambridge, MA.

Piattelli-Palmarini, M. (1994). Ever since language and learning: Afterthoughts on the Piaget-Chomsky debate. *Cognition*, **50**, 315–346.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, **44**, 547–555.

Pierrot-Desseilligny, C., Rivaud, S., Gaymard, B., Müri, R., and Vermersch, A.-I. (1995). Cortical control of saccades. *Annals of Neurology*, **37**, 557–567.

Pine, J. and Lieven, E. (1997). Slot and frame patterns in the development of the determiner category. *Applied Psycholinguistics*, **18**, 123–138.

Pinker, S. (1994). *The language Instinct: The New Science of Language and Mind*. Penguin Books.

Plate, T. (2003). *Holographic Reduced Representations*. CSLI Lecture Notes Number 150. CSLI Publications, Stanford, CA.

Plunkett, K., Sinha, C., Müller, M., and Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, **4**, 293–312.

Poliakoff, E., O'Boyle, D., Moore, P., McGlone, F., Cody, F., and Spence, C. (2003). Orienting of attention and Parkinson's disease: Tactile inhibition of return and response inhibition. *Brain*, **126**(9), 2081–2092.

Pollard, C. and Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Pollock, J.-Y. (1989). Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry*, **20**(3), 365–424.

Posner, M., Rafal, R., Choate, L., and Vaughn, J. (1984). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, **2**, 211–228.

Pouget, A. and Sejnowski, T. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, **9**(2), 222–237.

Pullum, G. and Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review*, **19**, 9–50.

Pulvermüller, F. and Asollahi, R. (2007). Grammar or serial order?: Discrete combinatorial brain mechanisms reflected by the syntactic mismatch negativity. *Journal of Cognitive Neuroscience*, **19**(6), 971–980.

Pulvermüller, F. and Knoblauch, A. (2009). Discrete combinatorial circuits emerging in neural networks: A mechanism for rules of grammar in the human brain. *Neural networks*, **22**(2), 161–172.

Pulvermüller, F., Lutzenberger, W., and Preissl, H. (1999). Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex*, **9**(5), 497–506.

Pulvermüller, F., Hauk, O., Nikulin, V., and Ilmoniemi, R. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, **21**(3), 793–797.

Pulvermüller, F (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, **5**, 517–524.

Pulvermüller, F (2003). *The Neuroscience of Language*. Cambridge University Press, Cambridge.

Pulvermüller, F (2010). Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain and Language*, **112**, 167–179.

Pylkkänen, L. (2002). *Introducing Arguments*. Ph.D. thesis, Dept of Linguistics and Philosophy, MIT.

Pylyshyn, Z. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, **80**, 127–158.

Pylyshyn, Z. and Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, **3**, 179–197.

Quiroga, R., Reddy, L., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, **435**, 1102–1107.

Radford, A. (2004). *Minimalist Syntax: Exploring the Structure of English*. Cambridge University Press, Cambridge, UK.

Rainer, G., Asaad, W., and Miller, E. (1998). Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(25), 15008–15013.

Rainer, G., Rao, C., and Miller, E. (1999). Prospective coding for objects in primate prefrontal cortex. *Journal of Neuroscience*, **19**(13), 5493–5505.

Ranganath, C. and Blumenfeld, R. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences*, **9**(8), 374–380.

Ranganath, C. and D'Esposito, M. (2005). Directing the mind's eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Current Opinion in Neurobiology*, **15**, 175–182.

Redington, M., Chater, N., and Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, **22**, 425–469.

Reynolds, J. and Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, **24**, 19–29.

Rhodes, B., Bullock, D., Verwey, W., Averbeck, B., and Page, M. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science*, **23**, 699–746.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, **2**, 1019–1025.

Rips, L. (1975). Quantification and semantic memory. *Cognitive Psychology*, **7**, 307–340.

Ritter, E. (1991). Two functional categories in noun phrases: Evidence from Modern Hebrew. In S. Rothstein, editor, *Syntax and Semantics 25: Perspectives in Modern Phrase Structure*, pages 37–62. Academic Press, New York.

Ritter, N., editor (2005). Special issue of *The Linguistic Review* (Vol. 22(2–4)) on 'The role of Linguistics in Cognitive Science'.

Rizzi, L. (1997). On the fine structure of the left-periphery. In L. Haegeman, editor, *Elements of Grammar*. Kluwer, Dordrecht.

Rizzolatti, G. and Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences*, **21**, 188–194.

Rizzolatti, G. and Matelli, M. (2003). Motor functions of the parietal lobe. *Experimental Brain Research*, **153**, 146–157.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996). Localization of grasp representations in humans by PET. 1. Observation versus execution. *Experimental Brain Research*, **111**(2), 246–252.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In M. Gazzaniga, editor, *The New Cognitive Neurosciences*, pages 539–552. MIT Press.

Ro, T., Farné, A., and Chang, E. (2003). Inhibition of return and the human frontal eye fields. *Experimental Brain Research*, **150**, 290–296.

Roberts, I. (2007). *Diachronic Syntax*. Oxford University Press, Oxford, UK.

Rohde, D. (2002). *A Connectionist Model of Sentence Comprehension and Production*. Ph.D. thesis, School of Computer Science, Carnegie Mellon University.

Rohrbacher, B. (1994). *The Germanic VO Languages and the Full Paradigm: A Theory of V to I Raising*. GLSA Publications, Amherst.

Rolls, E. (1996). A theory of hippocampal function in memory. *Hippocampus*, **6**, 601–620.

Rolls, E. (1999). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, **9**, 467–480.

Rolls, E., Robertson, R., and Georges-François, P. (1997a). Spatial view cells in the primate hippocampus. *European Journal of Neuroscience*, **9**, 1789–1794.

Rolls, E., Treves, A., Foster, D., and Perez-Vicente, C. (1997b). Simulation studies of the CA3 hippocampal subfield modelled as an attractor neural network. *Neural Networks*, **10**(9), 1559–1569.

Rolls, E., Aggelopoulos, N., and Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience*, **23**(1), 339–348.

Rolls, E., Grabenhorst, F., and Deco, G. (2010). Decision-making, errors, and confidence in the brain. *Journal of Neurophysiology*, **104**, 2359–2374.

Rosch, E. (1978). Principles of categorisation. In E. Rosch and B. Lloyd, editors, *Cognition and Categorisation*. Lawrence Erlbaum Associates, Hillsdale, N.J.

Rosch, E. and Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **8**, 382–439.

Rotman, G., Troje, N., Johansson, R., and Flanagan, R. (2006). Eye movements when observing predictable and unpredictable actions. *Journal of Neurophysiology*, **96**, 1358–1369.

Rotte, M., Koutstaal, W., Schacter, D., Wagner, A., Rosen, B., Dale, A., and Buckner, R. (2000). Left prefrontal activation correlates with levels of processing during verbal encoding: An event-related fMRI study. *NeuroImage*, **7**, S813.

Ruff, H. (1978). Infant recognition of invariant form of objects. *Child Development*, **49**(2), 293–306.

Rugg, M. (1995). ERP studies of memory. In M. Rugg and M. Coles, editors, *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. Oxford University Press, Oxford.

Rumelhart, D. and McClelland, J. (1995). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**(2), 159–188.

Rumelhart, D., McClelland, J., and the PDP research group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. MIT Press, Cambridge, MA.

Rushworth, M., Walton, M., Kennerley, S., and Bannerman, D. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, **8**(9), 410–417.

Sabes, P. (2000). The planning and control of reaching movements. *Current opinion in neurobiology*, **10**, 740–746.

Saffran, E. (2000). Aphasia and the relationship of language and brain. *Seminars in Neurology*, **20**(4), 409–418.

Saffran, J., Aslin, R., and Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928.

Sahin, N., Pinker, S., Cash, S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, **326**, 445–449.

Saito, N., Mushiake, H., Sakamoto, K., Itoyama, Y., and Tanji, J. (2005). Representation of immediate and final behavioral goals in the monkey prefrontal cortex during an instructed delay period. *Cerebral Cortex*, **15**, 1535–1546.

Sakata, H., Taira, M., Kunusoki, M., Murata, A., Tanaka, Y., and Tsustui, K. (1998). Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. *Philosophical Transactions of the Royal Society of London B*, **353**, 1363–1373.

Santorini, B. and Kroch, A. (2007). The syntax of natural language: An online introduction using the trees program. `http://www.ling.upenn.edu/ beatrice/syntax-textbook/ch7.html`.

Sapir, A., Soroker, N., Berger, A., and Henik, A. (1999). Inhibition of return in spatial attention: Direct evidence for collicular generation. *Nature Neuroscience*, **2**, 1053–1054.

Sapir, A., Hayes, A., Henik, A., Danziger, S., and Rafal, R. (2004). Parietal lobe lesions disrupt saccadic remapping of inhibitory location tagging. *Journal of Cognitive Neuroscience*, **16**(4), 503–509.

Sato, T. and Schall, J. (2003). Effects of stimulus-response compatibility on neural selection in frontal eye field. *Neuron*, **38**, 637–648.

Saygin, A., Wilson, S., Dronkers, N., and Bates, E. (2004). Action comprehension in aphasia: linguistic and non-linguistic deficits and their lesion correlates. *Neuropsychologia*, **42**(13), 1788–1804.

Scaife, M. and Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*, **253**, 253–254.

Schaal, S., Ijspeert, A., and Billard, A. (2003). Computational approaches to motor learning by imitation. *Philosophical Transactions of the Royal Society of London B*, **358**, 537–547.

Schabes, Y. (1990). *Mathematical and Computational Aspects of Lexicalized Grammars*. Ph.D. thesis, University of Pennsylvania.

Schall, J. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, **2**(1), 33–42.

Schank, R. and Abelson, R. (1977). *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Schieber, M. and Hibbard, L. (1993). How somatotopic is the motor cortex hand area? *Science*, **261**, 489–492.

Schlack, A., Sterbing-D'Angelo, S., Hartung, K., Hoffman, K., and Bremmer, F. (2005). Multisensory space representations in the macaque ventral intraparietal area. *Journal of Neuroscience*, **25**(18), 4616–4625.

Schmolesky, M., Wang, Y., Hanes, D., Thompson, K., Leutgeb, S., Schall, J., and Leventhal, A. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, **79**(6), 3272–3278.

Schneider, W. and Deubel, H. (2002). Selection-for-perception and selection-for-spatial-motor-action are coupled by visual attention: A review of recent findings and new evidence from stimulus-driven saccade control. In W. Prinz and B. Hommel, editors, *Attention and Performance XIX: Common Mechanisms in Perception and Action*, pages 609–627. Oxford University Press.

Schuller, A. and Rossion, B. (2004). Perception of static eye gaze direction facilitates subsequent early visual processing. *Clinical Neurophysiology*, **115**, 1161–1168.

Schultz, W., Dayan, P., and Montague, P. (1997). A neural substrate of prediction and reward. *Science*, **275**, 1593–1599.

Schwoebel, J. and Coslett, H. (2005). Evidence for multiple, distinct representations of the human body. *Journal of Cognitive Neuroscience*, **17**(4), 543–553.

Scoville, W. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery and psychiatry*, **20**, 11–21.

Seidenberg, M. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science*, **275**, 1599–1603.

Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, **47**, 277–287.

Servos, P., Osu, R., Santi, A., and Kawato, M. (2002). The neural substrates of biological motion perception: An fMRI study. *Cerebral Cortex*, **12**, 772–782.

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge University Press, Cambridge.

Shallice, T. and Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia*, **15**, 729–735.

Shallice, T., Rumiati, R., and Zadini, A. (2000). The selective impairment of the phonological output buffer. *Cognitive Neuropsychology*, **17**(6), 517–546.

Shapiro, K. and Caramazza, A. (2003). Grammatical processing of nouns and verbs in left frontal cortex? *Neuropsychologia*, **41**(9), 1189–1198.

Shapiro, K., Shelton, J., and Caramazza, A. (2000). Grammatical class in lexical production and morphological processing: Evidence from a case of fluent aphasia. *Cognitive Neuropsychology*, **17**(8), 665–682.

Shapiro, K., Pascual-Leone, A., Mottaghy, F., Gangitano, M., and Caramazza, A. (2001). Grammatical distinctions in the left frontal cortex. *Journal of Cognitive Neuroscience*, **13**(6), 713–720.

Shastri, L. (2001). Episodic memory trace formation in the hippocampal system: A model of cortico-hippocampal interaction. Technical Report TR-01-004, International Computer Science Institute (ICSI), UC Berkeley, Berkeley, CA.

Shastri, L. (2002). Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Sciences*, **6**(4), 162–168.

Shastri, L. and Ajjanagadde, V. (1993). From simple associations to systematic reasoning. *Behavioral and Brain Sciences*, **16**(3), 417–494.

Shibasaki, H. (1992). Movement-related cortical potentials. In A. Halliday, editor, *Clinical Neurology and Neurosurgery Monographs: Evoked Potentials in Clinical Testing (2nd Edition)*, pages 523–537. Churchill Livingstone, Edinburgh.

Shibasaki, H. and Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology*, **117**, 2341–2356.

Shibata, H., Kondo, S., and Naito, J. (2004). Organization of retrosplenial cortical projections to the anterior cingulate, motor, and prefrontal cortices in the rat. *49!!!!!!!!!!!!!!!*, pages 1–11.

Shikata, E., Hamzei, F., Glauche, V., Koch, M., Weiller, C., Binkofski, F., and Buchel, C. (2003). Functional properties and interaction of the anterior and posterior intraparietal areas in humans. *European Journal of Neuroscience*, **17**(5), 1105–1110.

Shima, K. and Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology*, **84**(4), 2148–2160.

Shima, K., Mushiake, H., Saito, N., and Tanji, J. (1996). Role for cells in the presupplementary motor area in updating motor plans. *Proceedings of the National Academy of Sciences of the United States of America*, **93**(16), 8694–8698.

Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature*, **445**, 315–318.

Shimamura, A., Janowsky, J., and Squire, L. (1990). Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia*, **28**, 803–813.

Shipp, S. (2004). The brain circuitry of attention. *Trends in Cognitive Sciences*, **8**(5), 223–230.

Siapas, A., Lubenov, E., and Wilson, M. (2005). Prefrontal phase locking to hippocampal theta oscillations. *Neuron*, **46**, 141–151.

Sibley, D., Kello, C., Plaut, D., and Elman, J. (2008). Large-scale modeling of wordform learning and representation. *Cognitive Science*, **32**, 741–754.

Siegel, J. (2008). *The Emergence of Pidgin and Creole Languages*. Oxford University Press, New York.

Silveri, M. and Misciagna, S. (2000). Language, memory and the cerebellum. *Journal of neurolinguistics*, **13**, 129–143.

Silveri, M., Leggio, M., and Molianri, M. (1994). The cerebellum contributes to linguistic production: A case of agrammatism of speech following right cerebellar lesion. *Neurology*, **44**, 2047–50.

Simons, J. and Spiers, H. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, **4**, 637–648.

Sincich, L. and Horton, J. (2005). The circuitry of V1 and V2: Integration of form, color and motion. *Annual Review of Neuroscience*, **28**, 303–326.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, **61**(1–2), 39–91.

Smith, D. and Mizumori, S. (2006). Hippocampal place cells, context, and episodic memory. *Hippocampus*, **16**(9), 716–729.

Snyder, L., Grieve, K., Brotchie, P., and Anderson, R. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature*, **394**, 887–891.

Snyder, L., Batista, A., and Anderson, R. (2000). Intention-related activity in the posterior parietal cortex: A review. *Vision Research*, **40**, 1433–1441.

Spelke, E., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, **97**(4), 605–632.

Spelke, E., Katz, G., Purcell, S., Ehrlich, S., and Breinlinger, K. (1994). Early knowledge of object motion: continuity and inertia. *Cognition*, **51**, 131–176.

Spelke, E., Kestenbaum, R., Simons, D., and Wein, D. (1995). Spatio-temporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, **13**, 113–142.

Spence, C. (2002). Multisensory attention and tactile information-processing. *Behavioural Brain Research*, **135**, 57–64.

Spence, S., Brooks, D., Hirsch, S., Liddle, P., Meehan, J., and Grasby, P. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain*, **120**, 1997–2011.

Sperber, R., McCauley, C., Ragain, R., and Weil, C. (1979). Semantic priming effects on picture and word processing. *Memory and Cognition*, **7**(5), 339–345.

Spiers, H. and Maguire, E. (2007). A navigational guidance system in the human brain. *Hippocampus*, **17**, 618–626.

Spivey, M., Tanenhaus, M., Eberhard, K., and Sedivy, J. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, **45**(4), 447–481.

Spratling, M. (2009). Learning posture invariant spatial representations through temporal correlations. *IEEE Transactions on Autonomous Mental Development*, **1**(4), 253–263.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, **2**, 447–457.

Starkey, P. and Cooper, R. (1980). Perception of numbers by human infants. *Science*, **210**, 1033–1035.

Steedman, M. (2000). *The Syntactic Process*. MIT Press/Bradford Books.

Stefan, K., Cohen, L., Duque, J., Mazzocchio, R., Celnik, P., Sawaki, L., Ungerleider, L., and Classen, J. (2005). Formation of a motor memory by action observation. *Journal of Neuroscience*, **25**(41), 9339–9346.

Stepankova, K., Fenton, A., Pastalkova, E., Kalina, M., and Bohbot, V. (2004). Object-location impairment in patient with thermal lesions to the right or left hippocampus. *Neuropsychologia*, **42**, 1017–1028.

Stowell, T. (1981). *The Origins of Phrase Structure*. Ph.D. thesis, MIT.

Stowell, T. (1996). The phrase structure of tense. In J. Rooryck and L. Zaring, editors, *Phrase Structure and the Lexicon*. Kluwer Academic Publishers, Dordrecht.

Stromswold, K., Caplan, D., Alpert, N., and Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, **52**(3), 452–473.

Sutton, R. and Barto, A. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel and J. Moore, editors, *Learning and computational neuroscience: Foundations of adaptive networks*, pages 497–537. MIT Press.

Suzuki, W. and Amaral, D. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *Comparative neuropsychology*, **350**(4), 497 – 533.

Szabolcsi, A. (1987). Functional categories in the noun phrase. In I. Kenesei, editor, *Approaches to Hungarian*, volume 2, pages 167–189.

Tabuchi, E., Mulder, A., and Wiener, S. (2003). Reward value invariant place responses and reward site associated activity in hippocampal neurons of behaving rats. *Hippocampus*, **13**, 117–132.

Taira, M., Mine, S., Georgopoulos, A., Murata, A., and Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Experimental Brain Research*, **83**, 29–36.

Talmy, L. (2000). *Towards a Cognitive Semantics. Volume 1: Concept Structuring Systems*. MIT Press.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, **19**, 103–139.

Tanaka, K. (1997). Mechanisms of visual object recognition: Monkey and human studies. *Current Opinion in Neurobiology*, **7**(4), 523–529.

Tanji, J. (2001). Sequential organization of multiple movements: Involvement of cortical motor areas. *Annual Review of Neuroscience*, **24**, 631–651.

Tanji, J., Shima, K., and Mushiake, H. (2007). Concept-based behavioral planning and the lateral prefrontal cortex. *Trends in Cognitive Sciences*, **11**(2), 528–534.

Taraldsen, T. (1990). D-projections and N-projections in Norwegian. In M. Nespor and J. Mascaró, editors, *Grammar in progress*. Foris, Dordrecht.

Tarski, A. (1936/1983). The concept of truth in formalized languages. In J. Corcoran, editor, *Logic, Semantics and Mathematics*. Hackett, Indianapolis.

Taube, J., Muller, R., and Ranck, J. (1990). Head direction cells recorded from the postsubiculum in freely moving rats. 1. Description and quantitative analysis. *Journal of Neuroscience*, **10**, 420–435.

Tettamanti, M., Buccino, G., Saccuman, M., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, **17**(2), 273–281.

Thompson, K. and Bichot, N. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, **147**, 251–262.

Thompson, K., Bichot, N., and Schall, J. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *Journal of Neurophysiology*, **77**, 1046–1050.

Thompson-Schill, S., D'Esposito, M., Aguirre, G., and Farah, M. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, **94**, 14792–14797.

Thornton, I., Rensink, R., and Shiffrar, M. (1999). Biological motion processing without attention. *Perception*, **28 (Suppl)**, 51.

Thornton, I., Cavanagh, P., and Labianca, A. (2000). The role of attention in the processing of biological motion. *Perception*, **29 (Suppl)**, 114.

Thráinsson, H (2001). Object shift and scrambling. In M. Baltin and C. Collins, editors, *The Handbook of Contemporary Syntactic Theory*, pages 148–202. Blackwell, Oxford.

Tipper, S. (1985). The negative priming effect: inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, **37A**, 571–590.

Tipper, S., Lortie, C., and Baylis, G. (1992). Selective reaching: Evidence for action-centered attention. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 891–905.

Tipper, S., Rafal, R., Reuter-Lorenz, P., Starrveldt, Y., Ro, T., Egly, R., Danzinger, S., and Weaver, B. (1997). Object-based facilitation and inhibition from visual orienting in the human split-brain. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 1522–1532.

Tipper, S., Howard, L., and Houghton, G. (1998). Action-based mechanisms of attention. *Philosophical Transactions of the Royal Society of London B*, **353**, 1385–1393.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Tomasello, M. (2005). Beyond formalities: The case of language acquisition. *Linguistic Review*, **22**(2–4), 183–197.

Tomasello, M. and Brooks, P. (1998). Young children's earliest transitive and intransitive constructions. *Cognitive Linguistics*, **9**, 379–395.

Tomasello, M., Call, J., and Gluckman, A. (1997). The comprehension of novel communicative signs by apes and human children. *Child Development*, **68**, 1067–1081.

Tomita, H., Ohbayashi, M., Nakahara, K., and Hasegawa, I Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, **401**, 699–703.

Tomlin, R. (1995). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts and E. Pederson, editors, *Language and Conceptualization*, pages 162–189. Cambridge University Press, Cambridge.

Toutanova, K., Manning, C., Flickinger, D., and Oepen, S. (2005). Stochastic HPSG parse disambiguation using the Redwoods corpus. *Research on Language and Computation*, **3**(1), 83–105.

Tranel, D., Adolphs, R., Damasio, H., and Damasio, A. (2001). A neural basis for the retrieval of words for actions. *Cognitive Neuropsychology*, **18**(7), 655–674.

Treiman, R. (1986). The division between onsets and rhymes in English syllables. *Journal of Memory and Language*, **25**, 476–491.

Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97–136.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, **24**(5), 295–300.

Trevena, J. and Miller, J. (2002). Cortical movement preparation before and after a conscious decision to move. *Consciousness and cognition*, **11**, 162–190.

Treves, A. (2005). Frontal latching networks: A possible neural basis for infinite recursion. *Cognitive Neuropsychology*, **22**(3–4), 276–291.

Treves, A. and Rolls, E. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus*, **2**, 189–200.

Trick, L. and Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? a limited capacity preattentive stage in vision. *Psychological Review*, **101**, 80–102.

Tsao, D., Vanduffel, W., Sasaki, Y., Fize, D., Knutsen, T., Mandeville, J., Wald, L., Dale, A., Rosen, B., Van Essen, D., Livingstone, M., Orban, G., and Tootell, R. (2003). Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron*, **39**, 555–568.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson, editors, *Organization of Memory*, pages 381–403. Academic Press, New York.

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford University Press, New York.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, **53**, 1–25.

Ullman, M. (2004). Contribution of memory circuits to language: The declarative/procedural model. *Cognition*, **92**, 231–270.

Ullman, M., Pancheva, R., Love, T., Yee, E., Swinney, D., and Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading and judgement of inflection in aphasia. *Brain and Language*, **93**, 185–238.

Ullman, S. (1984). Visual routines. *Cognition*, **18**, 97–159.

Umiltà, M., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. (2001). "I know what you are doing": A neurophysiological study. *Neuron*, **32**, 91–101.

Ungerleider, L. and Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. Goodale, and R. Mansfield, editors, *Analysis of Visual Behavior*. MIT Press.

Ungerleider, L., Courtney, S., and Haxby, J. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences*, **95**, 883–890.

Uno, Y., Kawato, M., and Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement—Minimum torque change model. *Biological Cybernetics*, **61**(2), 89–101.

Vallar, G. and Shallice, T. (1990). *Neuropsychological Impairments of Short-Term Memory*. Cambridge University Press, Cambridge, England.

van der Velde, F. and de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, **29**, 37–108.

Van Hoesen, G. (1982). The parahippocampal gyrus: New observations regarding its cortical connections in the monkey. *Trends in Neurosciences*, **5**, 345–350.

Vandenberghe, R., Mobre, A., and Price, C. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, **14**(4), 550–560.

Verfaillie, K., De Troy, A., and Van Rensbergen, J. (1994). Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **20**(3), 649–670.

Veselinova, L. (2008). Verbal number and suppletion. In M. Haspelmath, M. Dryer, D. Gil, and B. Comrie, editors, *The World Atlas of Language Structures Online*. Max Planck Digital Library, Munich. Chapter 39. Available online at `http://wals.info/feature/80`.

Vieville, T. (2006). About biologically plausible trajectory generators. In *Proceedings of the International Joint Conference on Neural Networks*, pages 563–572, Vancouver, Canada.

Vikner, S. (1995). *Verb movements and expletive subjects in Germanic languages*. Oxford University Press, New York.

von der Malsburg, C. and Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Biological Cybernetics*, **67**, 233–242.

Wagner, D., Maril, A., Bjork, R., and Schacter, D. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. *NeuroImage*, **14**, 1337–1347.

Wagner, D., Haxby, J., and Heatherton, T. (2012). The representation of self and person knowledge in the medial prefrontal cortex. *Wiley interdisciplinary reviews. Cognitive science*, **3**(4), 451–470.

Wallenstein, G., Eichenbaum, H., and Hasselmo, M. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences*, **21**, 317–323.

Walles, H. (2010). *A Perceptual Basis for Noun Phrase Syntax*. Ph.D. thesis, Department of Computer Science, University of Otago.

Walles, S., Knott, A., and Robins, A. (2008). A model of cardinality blindness in inferotemporal cortex. *Biological Cybernetics*, **98**(5), 427–437.

Walther, D., Itti, L., Riesenhuber, M., Poggio, T., and Koch, C. (2002). Attentional selection for object recognition: A gentle way. In S. W. Lee, H. Bülthoff, and T. Poggio, editors, *Biologically Motivated Computer Vision: Proceedings of the Second IEEE International Workshop*, pages 472–479. Springer, Berlin.

Webb, A., Knott, A., and MacAskill, M. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*, **133**, 51–56.

Weiss, S. and Rappelsburger, P. (2000). Long-range EEG synchronization during word encoding correlates with successful memory performance. *Cognitive Brain Research*, **9**, 299–312.

Werker, J. and Tees, R. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, **50**, 509–535.

Westermann, G. and Miranda, E. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language*, **89**, 393–400.

Wheeldon, L. and Levelt, W. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, **34**, 311–334.

Wheeler, M., Stuss, D., and Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin*, **121**(3), 331–354.

Whitlock, J., Sutherland, R., Witter, M., Moser, M.-B., and Moser, E. (2008). Navigating from hippocampus to parietal cortex. *Proceedings of the National Academy of the USA*, **105**(39), 14755–14762.

Wilson, B. and Baddeley, A. (1988). Semantic, episodic and autobiographical memory in a postmeningitic amnesic patient. *Brain and Cognition*, **8**(1), 31–46.

Wilson, F., Scalaidhe, S., and Goldman-Rakic, P. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, **260**(5116), 1955–1958.

Wilson, H. and Bergen, J. (1979). A four mechanism model for threshold spatial vision. *Vision Research*, **19**(1), 19–32.

Wilson, S., Saygin, A., Sereno, M., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, **7**(7), 701–702.

Wimmer, H. and Perner, J. (1983). Beliefs about beliefs—Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**(1), 103–128.

Wise, R., Scott, S., Blank, S., Mummery, C., Murphy, K., and Warburton, E. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain*, **124**, 83–95.

Wiskott, L. and Sejnowski, T. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, **14**, 715–770.

Wittgenstein, L. (1953/2001). *Philosophical Investigations*. Blackwell Publishers, Oxford, UK. Translated by G. Anscombe.

Wittman, B., Schott, B., Guderian, S., Frey, J., Heinze, H., and Düzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron*, **45**(3), 459–467.

Wolfe, J. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, **1**(2), 202–238.

Wood, E., Dudchenko, P., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, **397**, 613–616.

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, **69**(1), 1–34.

XTAG Research Group (2001). A lexicalized tree adjoining grammar for English. Technical Report IRCS-01-03, Institute for Research in Cognitive Science, University of Pennsylvania.

Xu, F. and Spelke, E. (2000). Large number discrimination in 6-month old infants. *Cognition*, **74**, B1f–B11.

Yimam, B. (2004). Agreement phenomena in Amharic. In V. Böll, editor, *Studia Aethiopica*, pages 319–336. Harrassovitz, Wiesbaden.

Zamparelli, R. (2000). *Layers in the Determiner Phrase*. Garland, New York.

Zoccolan, D., Cox, D., and DiCarlo, J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience*, **25**(36), 8150–8164.

Zurif, E., Caramazza, A., and Meyerson, R. (1972). Grammatical judgements of agrammatic aphasics. *Neuropsychologia*, **10**, 405–417.

Zwaan, R. and Taylor, L. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, **135**(1), 1–11.